



UNIVERSIDADE NOVA DE LISBOA
INSTITUTO DE HIGIENE E MEDICINA TROPICAL
MESTRADO EM CIÊNCIAS BIOMÉDICAS, especialidade
Biologia Molecular em Medicina Tropical e Internacional



Variabilidade dos genes que codificam proteínas de adesão às bactérias Gram-negativas em mosquitos vectores de malária.

Ana Sofia Laranjeira Lopes



**UNIVERSIDADE NOVA DE LISBOA
INSTITUTO DE HIGIENE E MEDICINA TROPICAL
MESTRADO EM CIÊNCIAS BIOMÉDICAS, especialidade
Biologia Molecular em Medicina Tropical e Internacional**



Variabilidade dos genes que codificam proteínas de adesão às bactérias Gram-negativas em mosquitos vectores de malária.

Ana Sofia Laranjeira Lopes

Tese apresentada para a obtenção do grau de Mestre em Ciências Biomédicas, especialidade de Biologia Molecular em Medicina Tropical e Internacional.

**Orientador:
Prof. Doutor Henrique Silveira**

**Co-Orientador:
Doutor João Pinto**

2008

Agradecimentos

Este trabalho foi realizado no Instituto de Higiene e Medicina Tropical (IHMT), no Centro de Malária e outras Doenças Tropicais (CMDT) e UEI Malária, para que tal se tivesse cumprido, não posso deixar de agradecer:

Ao Professor Doutor Henrique Silveira, pela orientação, apoio e disponibilidade, sempre demonstradas.

Ao Doutor João Pinto, pelo apoio e disponibilidade.

Ao coordenador do Centro de Malária e outras Doenças Tropicais, Professor Doutor Virgílio E. do Rosário e a toda a equipa, com um especial agradecimento à Cristina Mendes pela paciência, ajuda e apoio incondicional.

À Coordenadora do mestrado, Professora Doutora Maria Amélia Grácio.

À Directora de Serviço de Anatomia Patológica do Centro Hospitalar de Lisboa Ocidental (CHLO), Dr^a Maria Luísa Carneiro de Moura, pelo apoio e amizade.

Às minhas colegas do Serviço de Anatomia Patológica, pela paciência e ajuda. Um especial agradecimento à M^a José Pereira, Paula Laranjeiro e M^a Conceição Carvalho, acima de tudo pela amizade.

Aos amigos e família. Ao Renato e à Raquel.

Ao João. Ao meu irmão. Por tudo!

Aos meus pais, a quem tudo devo!

Índice Geral

1	Lista de abreviaturas e símbolos	8
2	Resumo	9
3	Introdução	11
3.1	Malária	11
3.1.1	O agente etiológico da malária	11
3.1.2	Manifestações da doença	13
3.2	O mosquito vector	14
3.2.1	Caracterização genética	15
3.2.2	Formas moleculares	16
3.3	Resposta imunológica <i>Anopheles gambiae</i>	17
3.3.1	Moléculas de reconhecimento – GGBP	18
3.3.2	GGBPB3	20
3.4	Evolução do sistema imunológico do mosquito	20
4	Objectivos	23
4.1	Objectivo geral	23
4.2	Objectivos específicos	23
5	Materiais e métodos	24
5.1	Proveniência e colecta da amostra	24
5.2	Identificação morfológica da espécie e forma molecular	25
5.3	Extracção de ADN	25
5.4	Reacção da polimerase em cadeia (PCR)	25
5.5	Condições de amplificação	26
5.6	Visualização do produto amplificado	27
5.7	Purificação dos produtos amplificados	27
5.8	Análise das sequências	27
5.8.1	Alinhamento das sequências	27
5.8.2	Polimorfismos e divergência	28
5.8.3	Testes de neutralidade	28
5.8.4	Diferenciação genética e diversidade molecular	29
5.8.5	Análise filogenética	29
6	Resultados	30
6.1	Polimorfismos e divergência	31

6.2	Testes de neutralidade.....	33
6.3	Diferenciação genética e diversidade molecular	38
6.4	Análise filogenética	41
7	Discussão.....	44
8	Conclusão	49
9	Referências bibliográficas	50
10	Anexos.....	60
	Anexo 1	61
	Anexo 2	62
	Anexo 3	63
	Anexo 4	64
	Anexo 5	65

Índice de figuras

Figura 1 - Ciclo de vida esporogónico e esquizogónico de <i>P. falciparum</i> spp. (http://www.dpd.cdc.gov/dpdx/HTML/Malaria.htm)	12
Figura 2 – Fêmea de <i>A. gambiae</i>	15
Figura 3 - Distribuição geográfica das amostras utilizadas.....	24
Figura 4 -Valores de Ka/Ks ao longo do gene GNBPB3	34
Figura 5 -Valores de Ka/Ks ao longo do gene GNBPB3	34
Figura6 - Valores de Ka/Ks ao longo do gene GNBPB3	34
Figura7- Valores de Ka/Ks ao longo do gene GNBPB3	35
Figura 8 - Valores de Ka/Ks ao longo do gene GNBPB3	35
Figura9 - Valores de Ka/Ks ao longo do gene GNBPB3	35
Figura10 - Valores de Ka/Ks ao longo do gene GNBPB3	35
Figura 11 -Esquema da sequência peptídica GNBPB3	36
Figura 12 - Rede de haplotipos referente às diferentes amostras das sete populações em estudo.	42
Figura13- Torso da rede de haplotipos referente às formas moleculares.....	42
Figura14- Torso da rede de haplotipos referente às formas moleculares e Ocidente <i>versus</i> Oriente	43
Figura 15- Rede de haplotipos referente às formas moleculares.....	64
Figura16- Rede de haplotipos referente às formas moleculares e Ocidente <i>versus</i> Oriente	65

Índice de tabelas

Tabela 1- Esquematização dos <i>Primers</i> utilizados no estudo.	26
Tabela 2 – Descrição dos programas de PCR.	26
Tabela 3 - Condições da mistura de reacção de PCR para GGBP27	27
Tabela 4- Número de amostras por região e respectiva forma molecular.....30	30
Tabela 5 - Numero de locais polimórficos e polimorfismos de ADN por região e forma (...)32	32
Tabela 6- Valores de Ka/Ks ao longo do gene GGBP3.....33	33
Tabela 7- Domínios da proteína GGBP3.....36	36
Tabela 8- Valores do teste de Tajima D, em intervalos de 75 nucleótidos.37	37
Tabela 9 - Intervalos de diferenciação genética38	38
Tabela 10- Diferenciação genética interpopulacional (divisão da amostra de acordo com a (...)). 39	39
Tabela 11- Diferenciação genética interpopulacional (divisão da amostra de acordo com (...)).39	39
Tabela 12- Diferenciação genética interpopulacional formas M versus formas S no Ocidente40	40
Tabela 13- Diferenciação genética interpopulacional40	40
Tabela 14-AMOVA comparando todas as populações e as diferentes regiões e formas (...)41	41
Tabela 15-Resultados do teste de Tajima D61	61
Tabela 16- Resultados do teste de Fu e Li.....62	62
Tabela 17- Resultados do teste de McDonald & Kreitman.....63	63

1 Lista de abreviaturas e símbolos

μl - microlitros

μM - micromole

ADN - ácido desoxirribonucleico

AMOVA – Análise da variância molecular

CdARN – ácido ribonucleico de cadeia dupla

CDC – *Centers for disease control and prevention*

CTL - lectinas do tipo C

DEF1 – defensina

dNTP's- Desoxirribonucleotídeos Fosfatados

FBN- domínio do fibrinogénio da família da imunoelectina

Fwd- *forward*

GNBP - proteínas de ligação a bactérias Gram-negativas

H - número de haplotipos

IHMT – Instituto de Higiene e Medicina Tropical

IMD – padrão de imunodeficiência

IUB - *International Union of Biochemistry*

IUPAC - *International Union of Pure and Applied Chemistry*

Ka – substituições não-sinónimas

KDR- resistência *knockdown*

Ks – substituições sinónimas

LRIM - proteínas do sistema imunológico ricas em leucina

M – concentração molar

MgCl₂ – cloreto de magnésio

N - nucleótidos

NOS –Oxido Nítrico sintetase

P - polimorfismo

PAMP- padrões moleculares associados aos patogénios

Pb – pares de bases

PCR – reacção da polimerase em cadeia

PGRP- proteínas de reconhecimento dos peptidoglicanos

PMS - péptidos anti microbianos

PRM – moléculas de reconhecimento de padrões associados aos patogénios

r.p.m.- rotações por minuto

REL2 – o gene ortólogo (de *Anopheles*) do gene Relish da *Drosophila*

Rev- reverse

SNP– polimorfismo de um único nucleótido

Taq polimerase – polimerase de ADN isolada da espécie *Thermus aquaticus*

TEP - proteínas que contêm tioesteres

UTR – região não traduzida

W- locais polimórficos

π- diversidade nucleotídica

2 Resumo

O mosquito *Anopheles gambiae* é o principal vector da malária humana em África, causada pelo parasita *Plasmodium falciparum*.

Avanços na manipulação genética de algumas espécies de insectos (Collins, 1994; Curtis, 1994), aliado à conclusão da sequência do genoma do *A. gambiae* na última década, permitiu entender os mecanismos que operam na relação mosquito-parasita, clarificando o papel do sistema imunológico inato como um potencial alvo no que se refere ao bloqueio no desenvolvimento do parasita (Osta *et al*, 2004).

O reconhecimento dos microrganismos pelo sistema imunológico é o primeiro passo da resposta à infecção. A família das GGBP que participam no reconhecimento e elaboração da resposta imunológica está representada em diversos grupos de insectos, sendo a GGBP3 de *A. gambiae* de especial interesse, já que parece ter um papel importante na defesa contra *P. falciparum* e *Plasmodium berghei* (Warr *et al.*, 2008) tornando-se um alvo de estudo importante.

Amplificou-se o gene que codifica para a proteína GGBP3, de amostras provenientes de *A. gambiae* s.s. de sete regiões da África Subsaariana. O gene foi posteriormente sequenciado, com o objectivo de caracterizar a sua variabilidade e evolução molecular, assim como detectar divergências populacionais relativamente às duas formas moleculares (M e S) e às diferentes regiões geográficas. Este estudo teve também como finalidade determinar o tipo de selecção a actuar e história evolutiva das populações em estudo.

Concluiu-se que todas as populações estão geneticamente separadas e que a distância entre as populações é provavelmente um factor determinante nas diferenças exercidas pela pressão selectiva, resultando assim, numa maior diferenciação entre as populações. Simultaneamente identifica-se que a forma molecular contribui menos para a diferenciação molecular das amostras em estudo, do que a sua localização geográfica. No que se refere ao tipo de selecção a actuar, os resultados indicam que o gene GGBP3 está sob a acção da selecção purificadora.

Summary

The mosquito *Anopheles gambiae* is the main vector of human malaria in Africa, caused by *Plasmodium falciparum*.

Advances in the genetic manipulation of some insect species (Collins, 1994; Curtis, 1994), allied to the conclusion of *A. gambiae* genome sequence allowed to understand the mechanisms operating at the mosquito-parasite interactions, clarifying the role of the *Anopheles* innate immune system as a potential way to block the development of the parasite (Osta *et al*, 2004).

Pathogen recognition is the first step of innate immune response. GGBP are a group of pattern recognition proteins largely represented in insects. Of special interest is the *A. gambiae* GGBP3 that might play an important role in the defence against *P. falciparum* and *Plasmodium berghei*. (Warr *et al.*, 2008), therefore becoming an important target of study.

The gene that codes for GGBP3 from *A. gambiae* s.s. of seven regions of Subsaharian Africa was sequenced with the purpose of characterizing its genetic variability, molecular evolution, population divergences and type of selection acting on its evolutionary history. Comparisons of the two molecular forms (M and S) and different geographic regions were performed.

All the populations studied are genetically separate and distance between the populations is probably a determinant factor in the differences observed as a result of selective pressure. Simultaneously molecular forms seem to contribute less for the molecular differentiation than the geographic localization. Data indicates that the gene coding for the GGBP3 is under purifying selection.

3 Introdução

3.1 Malária

A Malária é um problema urgente e real de saúde pública, pois, cerca de 40% da população mundial está em risco de contrair malária. Segundo a Organização Mundial de Saúde (World Malaria Report 2008), todos os anos mais de 500 milhões de pessoas são afectadas, causando a morte a mais de 1 milhão em todo o Mundo, sendo na sua maioria crianças na África subsaariana. No entanto, Ásia, América latina, Médio Oriente e algumas zonas da Europa são também afectadas.

A malária é causada por parasitas do género *Plasmodium* e transmite-se de pessoa para pessoa através da picada da fêmea de um mosquito infectado. O factor crucial na expansão e distribuição da doença é a existência de um vector, neste caso um mosquito pertencente ao género *Anopheles*. Dentro deste género existem várias espécies capazes de transmitir a doença, mostrando que o parasita se adaptou a diferentes espécies de mosquito, aumentando o potencial de disseminação da doença.

A transmissão da malária difere em intensidade dependendo da zona e das condições climáticas e proximidade de locais propícios ao desenvolvimento de mosquitos. Algumas regiões apresentam número constante de pessoas infectadas ao longo do ano, outras têm períodos do ano onde o número de infectados aumenta exponencialmente, coincidindo com a época das chuvas, denominada malária sazonal.

3.1.1 O agente etiológico da malária

O parasita da malária é um protozoário Apicomplexa do género *Plasmodium*. Existem mais de 100 espécies de *Plasmodium* que infectam diferentes animais desde os répteis até aos mamíferos. No entanto, na natureza apenas cinco espécies de *Plasmodium* infectam o Homem: *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale*, *Plasmodium malariae* e *Plasmodium knowlesi* (Greenwood *et al.*, 2008).

O ciclo de vida do *P. falciparum* é composto por duas fases, a sexuada exógena denominada de esporogónica (que ocorre no mosquito, hospedeiro definitivo) e a fase assexuada endógena ou esquizogónica (que ocorre no Homem, hospedeiro intermediário). Durante a refeição sanguínea a fêmea do mosquito inocula esporozoítos no hospedeiro humano¹. Os esporozoítos infectam os

hepatócitos² onde maturam e se diferenciam em esquizontes³, que quando maduros libertam merozoítos para a circulação sanguínea⁴. Esta fase denomina-se esquizogonia exo-eritrocitária^A. Nas espécies *P. vivax* e *P. ovale* podem persistir formas dormentes, os hipnozoítos, que permanecem nos hepatócitos e meses ou anos mais tarde podem iniciar a esquizogonia hepática. Após a replicação inicial no fígado, o parasita vai ter a sua multiplicação assexuada nas células sanguíneas (esquizogonia eritrocitária)^B. Os merozoítos infectam então os eritrócitos⁵ e diferenciam-se em trofozoítos que maturam em esquizontes. Quando maduros o eritrócito parasitado lisa, libertando merozoítos⁶ que irão invadir outros eritrócitos, perpetuando o ciclo (esta fase está associada ao síndrome febril). Alguns dos merozoítos que invadem os eritrócitos não evoluem para trofozoítos, mas diferenciam-se em gametócitos (formas sexuais)⁷. As formas sexuais masculinas denominam-se microgametócitos e as femininas macrogametócitos, que podem ser ingeridas pela fêmea Anophelina, durante a sua refeição sanguínea⁸. Inicia-se o ciclo esporogónico^C, no estômago do mosquito os gametócitos diferenciam-se em gâmetas masculinos e femininos que se fundem formando o zigoto⁹. O zigoto sofre alterações na sua morfologia, tornando-se alongado e móvel, designando-se oocineto¹⁰. Este vai invadir o epitélio do intestino médio do hospedeiro definitivo, diferenciando-se em oocisto¹¹. Dentro dos oocistos, o parasita multiplica-se e quando maduro, liberta esporozoítos¹² que vão invadir as glândulas salivares do mosquito. A inoculação dos esporozoítos no Homem perpetua o ciclo de vida da malária¹.

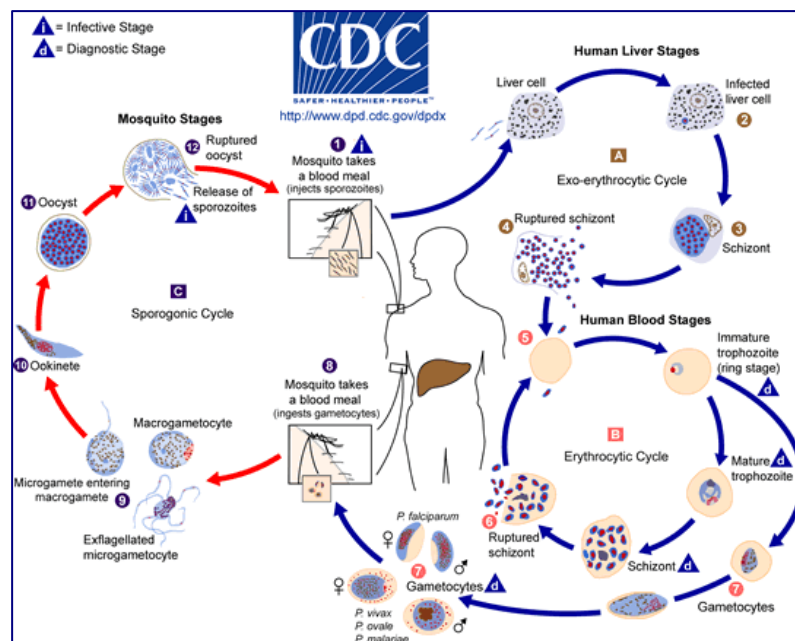


Figura 1 - Ciclo de vida esporogónico e esquizogónico de *P. falciparum* spp.
(<http://www.dpd.cdc.gov/dpdx/HTML/Malaria.htm>)

3.1.2 Manifestações da doença

Todas as espécies de *Plasmodium* que infectam o Homem apresentam semelhanças nos ciclos de transmissão, no entanto, estes ciclos dependem da patogenecidade do parasita e da velocidade a que se desenvolvem no organismo. Os sintomas mais comuns são febres recorrentes, arrepios, dores de cabeça, vômitos e suores, que ocorrem quando da ruptura dos eritrócitos e consequente invasão de novas formas parasitárias (Evans & Wellems, 2002). Diferentes espécies, causam picos febris em diferentes intervalos de tempo, por exemplo *P. falciparum*, *P. vivax*, e o *P. ovale* reinvadem os eritrócitos a cada 48 horas, enquanto que o *P. malariae* a cada 72 horas (Evans & Wellems, 2002).

O período de incubação da doença pode ir de 7 a 30 dias, sendo mais frequente os primeiros sintomas aparecerem nos 9 a 14 dias após inoculação do parasita. Os sintomas clássicos de malária são raramente observados, podendo durar os ataques entre 6 a 10 horas e ocorrer em cada dois ou em cada três dias (dependendo da espécie do parasita responsável pela infecção) e estão descritos em três estádios (*Centers for Disease Control and Prevention- CDC 2006, <http://www.cdc.gov/>*): calafrios, onde o paciente experimenta sensações de frio e tremores; calor, onde os sintomas são febre, dores de cabeça, vômitos e convulsões em crianças e por fim suores, a temperatura corporal volta ao normal e sensação de cansaço.

O diagnóstico clínico da doença é feito com base nos sintomas e sinais do paciente, que são inespecíficos, tornando-se necessário recorrer a testes laboratoriais para confirmação. Os mais frequentes são a observação microscópica de uma gota de sangue e testes imunológicos para pesquisa de antigénios.

O *P. falciparum* é a espécie que provoca a forma mais grave da doença, pois, altera a morfologia dos eritrócitos parasitados, que por seu lado adquirem capacidade de se ligarem às células endoteliais de pequenos vasos, causando a sua obstrução, designado, fenómeno de sequestro. Muitas vezes, estes capilares, são responsáveis pela microcirculação de órgãos vitais (pulmões, placenta e cérebro), que quando obstruídos, podem causar complicações graves ao indivíduo: Anemia grave, edema pulmonar e coma, podendo levar à morte (Evans & Wellems, 2002).

3.2 O mosquito vector

Os mosquitos do género *Anopheles* são os responsáveis pela transmissão da malária humana e são comuns na maioria dos países tropicais ou sub-tropicais. Existem cerca de 430 espécies de *Anopheles*, mas apenas trinta a quarenta são vectores na natureza. (Rey, 1992). Em África, os mosquitos pertencentes ao complexo de espécies *A. gambiae* são os vectores da malária mais importantes. A espécie *A. gambiae* pertence ao género *Anopheles*, Família Culicidae, Ordem Díptera, Classe Insecta, Filo Arthropoda do Reino Animalia (Rey, 1992). O complexo *A. gambiae* é constituído por sete espécies geneticamente aparentadas: *A. gambiae* (figura 2), *Anopheles arabiensis*, *Anopheles quadriannulatus* sp. A, *Anopheles quadriannulatus* sp. B, *Anopheles merus*, *Anopheles melas* e *Anopheles bwambiae*, que diferem na sua ecologia e comportamento (Parmakelis *et al.*, 2008)

Estes pequenos dípteros medem, em geral, menos de um centímetro de comprimento têm o corpo delgado e as patas longas. Os mosquitos adultos do género *Anopheles* distinguem-se dos outros mosquitos pelos seus palpos quase tão longos quanto as proboscides, pelas suas escamas brancas e pretas nas asas e pela postura, em repouso e quando se alimentam: cabeça para baixo, corpo em ângulo com as patas traseiras elevadas (Knell, 1991)

Grande parte dos anofelinos possui hábitos crepusculares e nocturnos. Durante o dia, dirigem-se para locais onde se abrigam da luz excessiva, do vento e dos predadores naturais. Nestes abrigos encontram também maior grau de humidade durante as horas mais quentes do dia. As diferentes espécies de anofelinos têm comportamentos distintos, como por exemplo, os hábitos alimentares, enquanto umas são antropofílicas, outras são zoofílicas; ou os locais de repouso e alimentação, sendo umas endofílicas e/ou endofágicas as outras exofílicas e ou exofágicas (Knell, 1991).

A longevidade dos anofelinos depende de factores genéticos próprios de cada espécie, denominada longevidade fisiológica e dentro de cada espécie as fêmeas possuem um tempo de vida, regra geral, superior à dos machos. Contribuem para a longevidade ainda os factores ecológicos, como a alimentação durante a fase larvar e adulta, as condições de humidade e temperatura ambiente e confronto com predadores (Rey, 1992). O seu tempo de vida é curto, no máximo de um mês, mas, na natureza, usualmente não ultrapassam as duas semanas. No que se refere às condições de temperatura e de humidade ideais, estas variam muito com as espécies e a literatura refere valores entre 20°C e 33°C, e 40% e 80% de humidade relativa (Rey, 1992).

Os anofelinos têm um ciclo de vida que passa por quatro estádios: ovo, larva, pupa e adulto. Os primeiros três estádios são aquáticos e duram entre 5 a 14 dias, dependendo da espécie em questão e da temperatura ambiente. É no estágio adulto que as fêmeas do mosquito *A. gambiae* são transmissoras da malária. As fêmeas desta espécie também se alimentam de néctares e fruta, mas para terem capacidade de colocarem ovos têm obrigatoriamente de fazer refeições sanguíneas. (Knell, 1991).



Figura 2 – Fêmea de *A. gambiae*

Foto de Anthony Cornel

(<http://entomology.ucdavis.edu/faculty/lanzaro/publications.htm>)

3.2.1 Caracterização genética

A conclusão da sequência do genoma do *A. gambiae*, juntamente com o sucesso na transformação das linhas germinais do mosquito (Grossman *et al.*, 2001) e a identificação de moléculas chave e da genética que afectam a susceptibilidade dos mosquitos à infecção por *Plasmodium*, em ambiente laboratorial (Barillas-Mury *et al.*, 1996; Blandin *et al.*, 2004) sugerem que o controlo da malária irá passar pela introdução de genes refractários em populações vectoras (Collins *et al.*, 2000).

A especificidade desta relação parasita-mosquito é demonstrada pela existência de anofelinos sensíveis e refractários à infecção por plasmódios (Brey *et al.*, 1995). Cada uma das etapas do ciclo de vida do parasita depende de interacções específicas entre estruturas moleculares do parasita e do hospedeiro. Em particular, o oocineto atravessa a membrana peritrófica e o epitélio do intestino, enquanto que os esporozoítos devem atravessar o epitélio das glândulas salivares. O resultado destas interacções é que dita a maior ou menor capacidade vectorial de cada anofelino para cada espécie de *Plasmodium*.

O *A. gambiae* possui 3 cromossomas, cromossoma 2 (2L; 2R), cromossoma 3 (3L;3R) e cromossoma X, num total de 278,253,050 pares de bases, documentando-se 1,106,441 SNP's. Existem 1,316 genes conhecidos que codificam para proteínas. (http://www.ensembl.org/Anopheles_gambiae/)

3.2.2 Formas moleculares

Populações de *A. gambiae sensu stricto* são altamente estruturadas em variadas formas, que são morfologicamente indistinguíveis. Inversões paracêntricas do braço direito do cromossoma 2 definem cinco diferentes citotipos ou formas cromossómicas: *Mopti*, *Bamako*, *Bissau*, *Forest*, e *Savanna*, e a variação da frequência destas formas correlaciona-se com as condições climáticas, da vegetação e mesmo com o próprio ambiente dos habitats humanos. (Holt *et al.*, 2002.).

Um sistema de classificação alternativo baseado nas diferenças fixas do ADN ribossomal, reconhece duas formas moleculares M e S. Estas formas foram identificadas pela primeira vez nos citotipos *Mopti* e *Savanna*, respectivamente. No entanto, análises feitas a populações de *A. gambiae* de diversas áreas do continente Africano mostra que nem sempre a forma molecular coincide com a forma cromossómica, o que pode ser explicado se assumir-se que os rearranjos cromossómicos não estão directamente envolvidos em mecanismos do isolamento reprodutivo, não especificando nenhuma unidade taxonomica diferente (Diabaté, *et al.*, 2008).

Segundo Coluzzi *et al.* (1985) *A. gambiae* continua em processo de especiação e as forças evolutivas, que contribuíram para a formação de subpopulações, não estão claramente identificadas contudo alterações ambientais provocadas pelo Homem no continente africano, contribuem em muito para a criação de novos nichos ecológicos, criando oportunidades para a especiação dos mosquitos. A selecção natural divergente, entre populações que habitam nichos ecológicos diferentes, é considerada a maior causa de especiação (Schluter, 1996). A adaptação ecológica associada ao processo de especiação, permitiu a exploração de ambientes, outrora excluídos, como as zonas áridas e as zonas sujeitas à irrigação para agricultura. Assume-se que a forma molecular M, adaptada ao clima árido, deslocou o habitat das suas formas larvares, de poças dependentes de chuvas (característica da forma molecular S) para zonas de irrigação agrícola (Diabaté, *et al.*, 2008).

Estudos referem que as duas formas moleculares são espécies incipientes (Lanzaro & Tripet, 2003), mas o isolamento reprodutivo não está completo, um exemplo disto é a presença do alelo *knock-*

down (kdr) que confere resistência aos insecticidas piretroides e é originário da forma S, passando presumivelmente desta para a forma M, por formação de um híbrido (Reimer *et al.*, 2005).

3.3 Resposta imunológica *Anopheles gambiae*

O sucesso no processo evolutivo dos insectos, considerado um dos grupos terrestres mais bem adaptados em termos de flexibilidade e eficiência na resposta imunológica, em muito se deve à grande diversidade de patógenos com que se encontram (Osta *et al.*, 2004).

Os insectos possuem mecanismos de resposta inata, tanto celular como humoral, altamente eficientes (Hoffmann *et al.*, 1996). Nos insectos a membrana peritrófica do intestino, a cutícula do exoesqueleto, e a linha da traqueia constituem barreiras físicas contra os invasores (Osta *et al.*, 2004). A membrana peritrófica é como uma luva protectora que circunda o estômago, composta por quitina, proteínas e proteoglicanos (Wang & Granados, 2001). Os oocinetos ingeridos de *P. falciparum* têm de penetrar a parede intestinal do mosquito *Anopheles* de modo a se desenvolverem em oocistos na camada basal do intestino. Este processo é facilitado pela acção das quitinases parasitárias que quebram as ligações da quitina existentes na membrana peritrófica (Langer *et al.*, 2000). Segundo um estudo elaborado por Shahabuddin *et al.* (1993), verifica-se que a invasão do intestino é totalmente bloqueada quando mosquitos são alimentados com sangue infectado contendo alosamidina (um inibidor da quitinase). Os microorganismos que conseguem ultrapassar estas barreiras físicas deparam-se com os mecanismos da resposta imunológica inata do hospedeiro tanto a nível celular, por parte dos hemócitos que correspondem às células sanguíneas do insecto, como a nível humoral, cujo responsável é o corpo gordo do mosquito (fisiologicamente equivalente ao fígado dos humanos) que liberta para a hemolinfa (o sangue do mosquito) factores de resposta imunológica que vão originar a fagocitose, secreção de péptidos antibacterianos, encapsulação aglutinação e melanização. (Osta *et al.*, 2004).

Alguns estudos (Tahar *et al.*, 2002; Lambrechts *et al.*, 2005) sugerem que a susceptibilidade de *Anopheles* ao *Plasmodium* depende do genotipo específico do vector e do parasita. Partindo deste pressuposto espera-se, que a nível molecular as interacções que ditam as relações vector-parasita deixem a sua assinatura nos genes envolvidos. Zheng *et al.*, (1997) sugere que muitas das incompatibilidades descritas têm bases genéticas simples. Uma alternativa às medidas de controlo tradicionais é interromper o desenvolvimento do *P. falciparum* no vector ou pelo menos diminuir de intensidade a sua transmissão. Esta abordagem necessita de um profundo conhecimento das

relações que ocorrem entre o parasita e o vector e dos mecanismos desta relação (Collins & Besansky, 1994; Collins & Paskewitz, 1995). O sistema imunológico, é assim, um importante candidato alvo para intervenções de controlo da malária (Cohuet *et al.*, 2008).

O sucesso no controlo da malária passa pelo manuseamento genético das populações de mosquitos, mais precisamente no que se refere ao controlo do ciclo de vida vector-parasita. É sabido que a grande maioria das combinações *Plasmodium* – mosquito são incompatíveis, muito devido às barreiras físicas e estruturais, mas também pelas barreiras do sistema imunológico (Dimopoulos, *et al.*, 2003). Só uma parte dos gametocitos de *Plasmodium* ingeridos pelo mosquito se desenvolvem em oocinetos e uma pequena parte destes se desenvolve em oocistos. Nos estádios mais tardios da infecção, apenas 20% dos esporozoítos invadem as glândulas salivares (Beier, 1998; Blandin & Levashina, 2004).

3.3.1 Moléculas de reconhecimento – GGBP

Os insectos são capazes de montar uma resposta à infecção por bactérias, fungos e parasitas. O reconhecimento dos microorganismos resulta na activação das células que irão desencadear mecanismos efectores como a síntese de péptidos antibacterianos, melanização ou encapsulação (Diabaté, *et al.*, 1993). Os mecanismos de defesa inata dos mosquitos, humorais e celulares são activados por moléculas de reconhecimento de padrões microbianos as proteínas de reconhecimento molecular (PRM's), altamente conservadas (Dziarski, 2004). Estas moléculas reconhecem e têm capacidade de se ligarem especificamente a padrões moleculares associados a patógenos, como por exemplo, lipopolissacaridos, presentes na parede das bactérias Gram-negativas (Michel & Kafatos, 2005) e peptidoglicanos, presentes especialmente na parede celular das bactérias Gram-positivas, mas também encontrados nas Gram-negativas (Dziarski, 2004).

Além da capacidade de reconhecimento estas moléculas podem promover directamente a morte celular dos patógenos por encapsulamento e fagocitose ou indirectamente pela activação de uma cascata de proteases serínicas e vias de sinalização intracelulares que controlam a transcrição de genes efectores (Warr, *et al.* 2000).

Duas destas vias de sinalização intracelulares foram caracterizadas em *Drosophila melanogaster*: a via *Toll* que é predominantemente activado por infecções fúngicas e bactérias Gram-positivas e a via de imunodeficiência (IMD), maioritariamente activado por bactéria Gram-negativas (Hoffmann *et al.*, 1996; Hoffmann & Reichhart, 2002).

Os estudos no sistema imunológico da *Drosophila* tiveram um grande impacto neste campo, pois, esta espécie tinha disponível uma grande variedade de ferramentas moleculares e genéticas, bem como, a conclusão da sequenciação do seu genoma. A partir do momento em que o genoma do *A. gambiae* foi sequenciado, e como um dos vetores mais importantes no que se refere à saúde humana, tornou-se rapidamente num modelo desejável de estudo (Holt *et al.*, 2002) especialmente com o estabelecimento de poderosas ferramentas na identificação da função dos genes que conduziu à identificação de cinco famílias de PRM's no genoma do mosquito, que se subdividem em: proteínas de reconhecimento dos peptidoglicanos (PGRP's) (Christophides *et al.*, 2002), proteínas de ligação a bactérias Gram-negativas (GNBP's) (Christophides *et al.*, 2002), proteínas contendo tioester (TEP's), proteínas ricas em leucina (LRIM's) e lectinas tipo C (CTL's) (Osta *et al.*, 2004). Estas moléculas presentes no sistema imunológico inato do mosquito, são um ponto focal de conflito entre o patógenos e o hospedeiro, que conduzem à sua evolução molecular.

A primeira GNBP a ser identificada foi no bicho-da-seda, espécie *Bombyx mori*, (a GNBP Bomo) onde se verificou que a proteína tinha a capacidade de se ligar a vários compostos presentes nas bactérias Gram-negativas (Lee *et al.*, 1996). Anos mais tarde, na mosca da fruta, *D. melanogaster* foram identificadas três famílias diferentes de GNBP's: DmGNBP-1, 2 e 3 (Kim *et al.*, 2000). A literatura refere que o padrão *Toll* na *D. melanogaster* é activado por padrões de reconhecimento molecular, sendo um deles da DmGNBP-1 (Kim *et al.*, 2000; Pili-Floury *et al.*, 2004). Esta proteína contém o domínio β -1,3 Glucanase que é homólogo ao existente na bactéria *Bacillus circulans* apenas faltando dois resíduos de ácido glutâmico, responsáveis pela actividade da glucanase.

A subfamília A das GNBP's pertence ao grupo todas as moscas da fruta e mariposas, conhecidas, enquanto que a subfamília B é específica dos mosquitos. O *A. gambiae* e a *D. melanogaster* partilham apenas um gene ortólogo que codifica, respectivamente, para as proteínas GNBPA2 e DmGNBP-3 (Christophides *et al.*, 2002).

Tendo em consideração o descrito, é de extrema importância entender o papel dos genes que codificam GNBP's, com especial ênfase para a GNBPA3, no sistema imunológico inato do mosquito *A. gambiae*, bem como as suas capacidades defensivas antibacterianas e especialmente anti-*Plasmodium*.

3.3.2 GNBPB3

O gene que codifica para a proteína GNBPB3 encontra-se no cromossoma 2, braço R, do *A. gambiae*. Apresenta na sua estrutura dois exões e um intrão, tendo o transcrito no total 1173 pares de bases e 390 resíduos de aminoácidos. O transcrito é o produto do gene AGAP002798 (*Ensemble identification gene*, em <http://www.ensembl.org>).

A proteína GNBPB3 pertence à família 16 das O-Glicosil-hidrolases, está relacionada com a resposta imunológica inata do mosquito e contém vários locais de glicosilação, característica de proteínas envolvidas na adesão ou reconhecimento célula-célula (Reuter & Gabius, 1999). As O-Glicosil-hidrolase são um vasto grupo de enzimas que hidrolisam ligações glicosídicas entre dois ou mais carboidratos e um não-carboidrato depiridometilado. O sistema de classificação para as Glicosil-hidrolases, baseado na similaridade, levou à definição de 85 famílias diferentes (<http://www.cazy.org/>). As O-Glicosil-hidrolase da família 16 compreendem uma série de actividades, entre elas: *lichenase*; *xyloglucan xyloglucosyltransferase*; *agarase*; *kappa-carrageenase*; *endo-beta-1,3-glucanase*; *endo-beta-1,3-1,4-glucanase* e *endo-beta-galactosidase* e tem como função a actividade hidrolítica e capacidade hidrolítica de compostos O-glicosil. (<http://smart.embl-heidelberg.de/> número de acesso PF00722).

Verificou-se que após o silenciamento dos genes que codificam para várias GNBP, o silenciamento da GNBPB3 têm um forte impacto na permissividade à infecção por *P. berghei*. O mesmo estudo, indica que as *GNBPBA2* e *B3* terão provavelmente um papel importante na defesa contra as *P. falciparum* e *P. berghei* e *Escherichia coli*. (Warr *et al.*, 2008).

3.4 Evolução do sistema imunológico do mosquito

O sistema imunológico inato para ser efectivo tem de ter a capacidade de reconhecer os patógenos invasores, activar o sinal e destruir o agente causador da infecção. O patógeno por outro lado, tem de ter a capacidade de se evadir, de modo a sobreviver ao sistema imunológico do hospedeiro (Lazzaro, 2008). Assim sendo, o sistema imunológico do hospedeiro deve idealmente evoluir no sentido de se manter efectivo, no que respeita a alterações na natureza dos organismos patogénicos a que está exposto e também à evolução dos mecanismos de virulência, por parte dos agentes infecciosos (Sackton, 2003).

Como já referido, a defesa humoral da *Drosophila* é activada através de proteínas de reconhecimento molecular (PRM's) (Osta *et al.*, 2004), levando à segregação de péptidos anti microbianos (PMS's). Seria então de esperar que tanto os PMS's como as PRM's experimentem pressão da selecção natural, pois, entram em contacto directo com os microrganismos invasores. No entanto, não é isso que se tem verificado: as proteínas de sinalização envolvem rapidamente enquanto que as PMS's e as PRM's mostram pouca indicação de adaptação ao nível da sequência de aminoácidos (Lazzaro, 2008).

As interacções moleculares e celulares são um pré-requisito para um sistema parasita-vector tornar-se estável e co-evoluir (Alavi *et al.*, 2003). A existência na natureza de combinações parasita-vector pouco eficazes, como por exemplo: *Plasmodium gallinaceum*/ *Anopheles stephensi* ou mesmo incompatíveis como *P. berghei*/ *Aedes aegypti*), sugerem que os parasitas co-evolvem com os seus hospedeiros naturais. Reforçando o descrito anteriormente em 2002, Tahar *et al.* descreve que a invasão do epitélio do intestino médio de *A. gambiae* por *P. falciparum* não promove a mesma sobrerregulação de marcadores imunológicos quando infectado com *P. berghei*.

Até há pouco tempo, os mecanismos moleculares que controlam o número de parasitas no vector permaneciam incompreensíveis, podendo apenas ser descritos com base na analogia da imunidade antibacteriana por estudos efectuados *in vitro*. O despoletar de técnicas que envolvem o silenciamento de genes por mediação de dcARN, *in vivo*, veio mudar a situação radicalmente, permitindo a identificação de inúmeros produtos génicos no mosquito que actuam como antagonistas ou agonistas no desenvolvimento do parasita (Levashina *et al.*, 2001). Em 2002, Blandin *et al.*, (2004) descreve que quando inserido dsRNA correspondente ao gene da defensina (DEF 1) de *A. gambiae* na cavidade corporal de mosquitos adultos o desenvolvimento de *P. berghei* não é afectado. Recentemente foram identificados produtos genicos que actuam como antagonistas TEP1 (proteína contendo tioester) e LRIM1 (proteína rica em leucina). A TEP1 do *A. gambiae* medeia o reconhecimento e destrói os oocinetos do parasita *P. berghei*. Foram identificados dois alelos da TEP1, a TEP1s (susceptível) e a TEP1r (refractária). Estas duas formas codificam para proteínas 93% idênticas e com 96% de similaridade (sem inserções e deleções), no entanto, mostram diferenças significativas no que se refere à sua ligação aos oocinetos. A expressão da TEP1s lisa 80% dos oocinetos de *P. berghei*, depois de estes atravessarem o epitélio do intestino, ainda assim, um número substancial consegue formar oocistos e continuar o ciclo de vida. Pelo contrário, aquando a expressão da TEP1r os parasitas são destruídos por lise e a sua eliminação é facilitada por melanização (Baxter *et al.*, 2007).

Num estudo recente, realizado por Parmakelis *et al.* (2008), foram analisados quatro genes (gambicina, NOS, Rel2 e FBN9) envolvidos na resposta imunológica das diferentes espécies do complexo *A. gambiae*, que se suspeitava estarem sob pressão selectiva, imposta pelos parasitas da malária. Os autores concluíram que os genes estudados interagem com o *Plasmodium*, independentemente de se envolverem com outros agentes microbianos, apresentando uma resposta adaptativa à infecção pelo parasita. Verificaram que os genes em estudo apresentam sinais de introgressão e/ou de polimorfismos ancestrais partilhados, não se observando, no entanto, sinais de selecção positiva em nenhum dos genes, levando a crer que a aquisição do *P. falciparum* por cada uma das populações ou espécies são eventos evolucionários independentes. A gambicina é um importante péptido antibacteriano e letal para o *P. berghei*; O FB9 encontra-se sobre-regulado em invasões de *P. falciparum* em *A. gambiae* e o Rel2 regula a expressão de vários genes antibacterianos em *A. gambiae*, entre eles antagonistas do parasita da malária (como o TEP1 e o LRIM1), no entanto, por estarem envolvidos em diversos eventos antimicrobianos, reduz as hipóteses de direccionar uma resposta imunológica específica anti-*Plasmodium*.

As famílias das PGRP's e das GGBP's experimentaram um fluxo considerável nas linhagens mais antigas que separam as moscas, mosquitos, abelhas e besouros com numerosas expansões e contrações nos tamanhos das famílias (Sackton *et al.*, 2007). As PRM's são, do ponto de vista evolutivo, mais estáveis nos insectos que contêm PGRP's e GGBP's (pelo menos entre aqueles que têm o genoma sequenciado). Este facto é indicativo de algum rearranjo genómico relativamente aos genes da *Drosophila* (Lazzaro *et al.*, 2007) ou mesmo entre os genes do *Aedes aegypti* e *A. gambiae* (Waterhouse R.M. *et al.*, 2007). Sugerindo que as flutuações genómicas experimentadas pelos diferentes insectos se relacionem com os diferentes tipos de patógenos que cada um deles encontra.

Se o sistema imunológico dos insectos é um potencial aliado no controlo da malária, então torna-se essencial e relevante assinalar um ou vários genes, com relevância no sistema imunológico do insecto e que sejam fáceis de manipular. O conhecimento de que as GGBP fazem parte e participam na resposta imunológica dos insectos, com especial interesse na GGBPB3 que têm um papel importante na defesa contra as *P. falciparum* e *P. berghei* e *E. coli*. (Warr *et al.*, 2008) torna-a um alvo de estudo importante. É, assim, de todo interesse aprofundar o conhecimento deste gene, no que diz respeito à sua variabilidade genética e evolução molecular, em populações de *A. gambiae* s.s. em regiões da África subsaariana, onde a malária é endémica. Este conhecimento é essencial já que a substituição de mosquitos selvagens por mosquitos geneticamente modificados, refractários à infecção por *P. falciparum*, surge cada vez mais como uma estratégia para controlar a transmissão da malária (Parmakelis *et al.*, 2008).

4 Objectivos

4.1 Objectivo geral

Caracterizar a variabilidade e a evolução molecular do gene que codifica para a proteína GNBPB3, nas duas formas moleculares de *A. gambiae* s.s. em diferentes zonas geográficas da África subsaariana.

4.2 Objectivos específicos

- Caracterizar e comparar a diversidade genética em diferentes zonas geográficas;
- Caracterizar a estrutura genética das populações de *A. gambiae* s.s.;
- Determinar o tipo de selecção e história evolutiva das populações em estudo;
- Caracterizar e comparar a diversidade genética nas diferentes formas moleculares M e S;
- Identificar possíveis divergências populacionais nas formas moleculares M e S.

5 Materiais e métodos

5.1 Proveniência e colecta da amostra

A amostra tem no total 111 mosquitos da espécie *A. gambiae* s.s. capturados em sete regiões da África subsaariana (figura 3) e distribuem-se do seguinte modo: Angola-Luanda com 7 amostras; Gabão-Dienga com 14 amostras; Ghana-Accra com 15 amostras; Ghana-Okyereko com 22 amostras; Guiné-Bissau com 26 amostras; Moçambique com 15 amostras e Tanzânia com 12 amostras.



Figura 3 - Distribuição geográfica das amostras utilizadas.

As circunferências vermelhas indicam o país de recolha de mosquitos.

Mapa do continente africano retirado de <http://paginas.terra.com.br/arte/mundoantigo/africa/>

A recolha dos *A. gambiae* no Oeste Africano foi feita em Moçambique e na Tanzânia. Em Moçambique as amostras foram recolhidas em Furvela (província do Inhambane) entre Fevereiro e Abril de 2004 utilizando a técnica de armadilha luminosa. As amostras da Tanzânia foram recolhidas na cidade de Ifakara em 2003, também com armadilha luminosa em cada duas semanas, no período de um ano (Drakeley *et al.*, 2003). Em Angola, as amostras foram recolhidas em Luanda, utilizando a técnica “captura em repouso”, em 2003. No Ghana, tanto em Accra como em Okyereko, e na Guiné-Bissau, as amostras foram também recolhidas por “captura em repouso”, em 2006 e 1996, respectivamente. Por fim, no Gabão-Dienga, a colheita foi feita sobre isca humana em 1999.

Os mosquitos foram guardados em tubos individuais com sílica gel e transportados à temperatura ambiente para o Instituto de Higiene e Medicina Tropical (IHMT) em Lisboa, onde foram conservadas a -20°C até à extracção de ADN.

5.2 Identificação morfológica da espécie e forma molecular

A identificação do género e da forma sexual do *Anopheles* foi baseada nas suas características morfológicas pela visualização directa ao microscópio óptico. A confirmação da espécie *A.gambiae* s.s. e das formas moleculares foi realizada segundo a técnica descrita em Fanello *et al.*, 2002.

5.3 Extracção de ADN

A extracção de ADN foi feita segundo o método descrito por Collins *et al.* em 1988: os mosquitos adultos foram macerados em tubos *ependorfs* de 1,5 ml onde se acrescentou 25 µl de tampão de lise de modo a inibir a actividade das *DNases*, as amostras foram incubadas a 68°C, durante 30 minutos. De seguida adicionou-se 7 µl de Acetato de Potássio 8M, para a precipitação das proteínas, e incubou-se durante 30 minutos num banho de gelo. As amostras foram posteriormente centrifugadas a 12000 rpm (rotações por minuto) durante 10 minutos. O sobrenadante foi transferido para um novo tubo onde se adicionou 100 µl de etanol a 95%, de modo a precipitar o ADN; centrifugou-se novamente a 12000 rpm, durante 10 minutos, desperdiçou-se o sobrenadante. O depósito (ADN) ficou a repousar durante a noite a uma temperatura de -20°C. No dia seguinte, o ADN foi lavado com etanol a 70%, seco ao ar e ressuspendido em TE e conservado a -20°C.

5.4 Reacção da polimerase em cadeia (PCR)

Os *primers* utilizados para amplificar a totalidade do gene da GNBPB3 (1092 pb) foram desenhados com base nas sequências do genoma do *A. gambiae* s.s. depositadas no *Ensemble* (<http://www.ensemble.org>). Para o primeiro *nested* PCR foi desenhado um par de *primers*, a montante e a jusante do gene (tabela 1). Uma vez que o produto da reacção anterior era demasiado grande para sequenciar, foram efectuadas 3 reacções subsequentes em *nested* PCR, correspondendo à extremidade 3', ao centro e à extremidade 5' da sequência (tabela 1).

Tabela 1- Esquematização dos *Primers* utilizados no estudo.

<i>Tamanho do fragmento</i>	<i>Nome do primer</i>	<i>Sequência do primer 5'-3'</i>	<i>Temperatura de annealing</i>
1902 pb	GNBPB3-out-fwd	CACTCCAGCGAACATTTGTG	51,4 °C
	GNBPB3-out-rev	CTTCAGTGTGTGGCGGTTTA	
851 pb	GNBPB3-in-3-fwd	CCCTAAATAAAGCGGCACAC	
	GNBPB3-in-3-rev	GCACTCTTGATGGGGTTGAT	
963 pb	GNBPB3-in-5-fwd	GTTCTGGGGATGTGAGCGTA	
	GNBPB3-in-5-rev	CAGGGATCTTTTGCGTGATT	
571 pb	GNBPB3-centro-fwd	ACRGGAGAGCTGATCTTTGA	55°C
	GNBPB3-centro -rev	GCCWCGRTAGTCCATATTGC	

5.5 Condições de amplificação

As PCRs foram efectuadas no Termociclador T1 (Biometer, Alemanha). A reacção de PCR consistiu numa primeira etapa onde ocorre um primeiro ciclo de desnaturação da molécula de ADN a 95° durante 2 minutos, seguida de 35 ciclos de amplificação que consistem em: 1 minuto a 95°C para desnaturação da molécula de ADN, 30 segundos a 51,4°C de *annealing* (promove a ligação do ADN, neste caso, dos *primers* à molécula alvo) e um período de extensão a 72°C a 1min, e uma extensão final de 5 minutos (tabela 2).

Tabela 2 – Descrição dos programas de PCR.

<i>Programa primers "out"</i>	<i>Programa primers "3in" e "5 in"</i>	<i>Programa primers "centro"</i>
Passo 1 – 95°C a 2 minutos	Passo 1 – 95°C a 2 minutos	Passo 1 – 95°C a 2 minutos
35x	35x	35x
Passo 2- 95°C a 1 minuto	Passo 2- 95°C a 1 minuto	Passo 2- 95°C a 30 minutos
Passo 3- 51,4°C a 30 segundos	Passo 3- 51,4°C a 30 segundos	Passo 3- 55°C a 30 segundos
Passo 4- 72°C a 2 minutos	Passo 4- 72°C a 1 minuto	Passo 4- 72°C a 1 minuto
Passo 5- 72°C a 5 minutos	Passo 5- 72°C a 5 minutos	Passo 5- 72°C a 5 minutos
Passo 6 – 4°C	Passo 6 – 4°C	Passo 6 – 4°C

As misturas de reacção utilizada para cada uma das etapas anteriormente descritas foram efectuadas de acordo com a seguinte tabela (tabela 3):

Tabela 3 - Condições da mistura de reacção de PCR para GNB

		Primers "out"	Primers "5in" e "3 in"	Primers "centro"
	stock	Amostra (µl)	Amostra (µl)	Amostra (µl)
Água		29,5	29,5	33,5
Tampão	5x	10	10	10
MgCl ₂	50µM	3	3	3
dNTP's	200 µM	1	1	1
Primer F	10µM	2,5	2,5	0,5
Primer R	10µM	2,5	2,5	0,5
Taq	5U/µl	0,5	0,5	0,5
Volume total		49	49	49

5.6 Visualização do produto amplificado

Para confirmar a amplificação, procedeu-se à visualização dos produtos da reacção em gel de agarose a 2%, em tampão TBE (0,1 M Tris, 5mM EDTA e 0,1 M Ácido Bórico, pH8,3), corada com brometo de etídio. A electroforese foi efectuada durante 1 hora a 140 Volts. O gel foi visualizado no transiluminador UVIDOC (Sigma, Estados Unidos da América)

5.7 Purificação dos produtos amplificados

Todos os produtos de PCR, antes de serem enviadas para sequenciação, foram purificados com o *kit* SureClean (BIOLINE, Reino Unido) segundo o protocolo descrito pelo fabricante. As amostras foram enviadas, para sequenciar, para o laboratório da Macrogen (Coreia).

5.8 Análise das sequências

5.8.1 Alinhamento das sequências

Após análise da qualidade das sequências nos fluorogramas, todas as sequências foram corrigidas no programa informático *BioEdit Sequence Alignment Editor 7.0.5.2*. (Hall T.A., 1999), de modo a obter a sequência consenso para cada uma das amostras. Verificou-se que ao longo das sequências, na maioria das amostras, existiam ambiguidades entre dois nucleotidos (N). Com o auxílio dos electrofluorogramas (no *software* Chromas) tentou identificar-se o nucleotido correcto, que na

maioria não foi possível, pois, correspondiam a locais heterozigóticos, tendo sido corrigidas de acordo com a tabela IUB/ IUPAC.

5.8.2 Polimorfismos e divergência

Neste estudo, a análise do número de polimorfismos (P), a diversidade haplotípica (h) e a diversidade nucleotídica (π) foram determinados para todas as sequências no programa DNAsp 4.50.1 (Rozas *et al.*, 2003)

5.8.3 Testes de neutralidade

Foram utilizados vários testes que assentam no modelo neutro da evolução, que partem do pressuposto que a população em estudo tem tamanho constante, acasalamento ao acaso e não é estruturada. Estes testes são utilizados para identificar os desvios à neutralidade e o tipo de selecção a actuar. De acordo com a teoria da neutralidade a maioria das mutações que se fixam numa população é neutra, e como tal não influenciam o *fitness* do indivíduo ou a sobrevivência da população (Kimura, 1968).

Neste estudo realizaram-se os testes de Tajima D (Tajima, 1989) e de Fu e Li (Fu & Li, 1993), que se baseiam na comparação da frequência e tipo de mutações numa população, que se espera que sejam neutras.

O número de substituições não sinónimas (K_a) e sinónimas (K_s), entre sequências de ADN pode ser utilizado como indicador da pressão selectiva a actuar num gene que codifica para uma proteína. Assim, espera-se que o rácio K_a/K_s seja igual a zero quando não há desvios à neutralidade, e que seja superior a um, nos casos de selecção positiva e inferior a um, nos casos de selecção purificadora.

Utilizou-se também o teste de McDonald-Kreitman, que revela o rácio entre os locais polimórficos sinónimos e não-sinónimos entre populações e as mutações fixas sinónimas e não-sinónimas, numa tabela de contingência 2x2. Os valores de p são calculados usando o teste de Fisher (McDonald & Kreitman, 1991). No caso de, estarmos perante a acção da selecção positiva é esperado um aumento do número de substituições não-sinónimas, comparativamente a sinónimas.

Todos os testes de neutralidade foram realizados no *software* DNAsp 4.50.1 (Rozas *et al.*, 2003)

5.8.4 Diferenciação genética e diversidade molecular

A caracterização e comparação da diferenciação (estruturação) genética entre as diferentes populações em estudo foram feitas utilizando o índice de fixação (F_{st}) (Wright, 1978) que mede a proporção da diferenciação genética total distribuída entre as populações em estudo. A variação molecular entre espécies foi feita através da AMOVA (*Analysis of Molecular Variance*). A AMOVA permite-nos estudar a existência de populações diferenciadas bem como, o grau de significância da variabilidade genética intra e interpopulacional. (Excoffier *et al.*, 2006). Baseia-se na análise de diferentes níveis hierárquicos, onde o primeiro (F_{ct}) é dado pela permutação entre os grupos; o segundo (F_{sc}) entre as populações dentro dos grupos e o terceiro (F_{st}) dentro das populações.

Os dois testes foram efectuados no programa informático Arlequin 3.1.1. (Schneider *et al.*, 2000).

5.8.5 Análise filogenética

A genealogia entre os haplotipos foi conseguida através da construção de uma rede de haplotipos elaborada no programa NETWORK 4.5.0.7 (*Phylogenetic network constructions*). Este programa estima a relação entre haplotipos com base na parcimónia, agrupando-os pelas mutações obtidas nas suas sequências (Röhl, 2003).

6 Resultados

As 111 amostras de mosquitos distribuem-se segundo o local e a forma molecular do seguinte modo: Angola-Luanda com 7 amostras, (forma molecular M); Gabão - Dienga com 14 amostras (forma molecular S); Ghana-Accra com 15 amostras (forma molecular S); Ghana-Okyereko com 22 amostras (12 correspondentes à forma molecular S e 10 à forma molecular M); Guiné-Bissau com 26 amostras (11 correspondentes à forma molecular S e 15 à forma molecular M); Moçambique com 15 amostras (forma molecular S) e Tanzânia com 12 amostras (forma molecular S). Destas 111 amostras iniciais, apenas 70 entraram no nosso estudo (tabela 4, uma vez que as restantes 41 amostras foram eliminadas por não ter sido possível fazer a sua amplificação ou pela má qualidade das sequências obtidas.

Obteve-se então, 70 sequências, com 1335 nucleotidos (N) cada, que inclui o 3' UTR (1 ao 91 N), Exão 1 (92 ao 301 N), Intrão (302 ao 371 N), Exão 2 (372 ao 1324 N) e o 5' UTR (1325 ao 1335 N).

Tabela 4- Número de amostras por região e respectiva forma molecular

<i>Localização</i>	Nº de amostras final
Angola-Luanda (M)	6
Gabão-Dienga (S)	9
Ghana-Accra (S)	7
Ghana-Okyereko (S)	6
Ghana-Okyereko (M)	6
Guiné-Bissau (S)	7
Guiné-Bissau (M)	12
Moçambique (S)	10
Tanzânia (S)	7

(M) - forma molecular M; (S) - forma molecular S

6.1 Polimorfismos e divergência

A análise dos polimorfismos (tabela 5) ao longo da sequência mostra que a região que apresenta o maior número de locais polimorficos é a Guiné-Bissau com um total de 76 na zona codificante, correspondendo 68 ao exão 2, seguindo-se o Ghana – Okyereko com 58, correspondendo 50 ao exão 2. No que se refere ao número de haplotipos distintos em cada amostra (h), Guiné-Bissau e Ghana – Okyereko têm também os valores mais altos, sendo 19 e 12 respectivamente.

A diversidade genética (tabela 5), tanto baseada na diversidade nucleotídica (π) como baseada em locais polimorficos (w) apresenta valores muito baixos em todas as amostras analisadas, documentando-se os mais elevados nas regiões descritas anteriormente Guiné-Bissau ($\pi = 0,014$; $w = 0,018$) e Ghana-Okyereko ($\pi = 0,016$; $w = 0,016$).

Relativamente às formas moleculares, observa-se que a forma molecular M apresenta um π e W (0,015 e 0,017 respectivamente) idêntico à forma molecular S (0,014 e 0,017 respectivamente), se tivermos em conta o número de indivíduos presente em cada amostra, verifica-se que a forma molecular M tem um maior número de polimorfismos ($P = 75$ para $n = 24$) do que a forma molecular S ($P = 88$ para $n = 46$). No que se refere ao número de haplotipos distintos em cada amostra, observa-se que para a forma molecular S, $h = 39$ e para a forma molecular M, $h = 23$. Curiosamente observam-se valores sensivelmente superiores na zona não codificante (intrão) do que na zona codificante (exão 1 e exão 2) (tabela 5).

No estudo efectuado ao longo do gene identifica-se a presença de polimorfismos, sendo as populações com o número mais elevado Ghana-Okyereko e Guiné-Bissau (58 e 76, respectivamente) e as que apresentaram o menor número de locais polimorficos Angola e Moçambique (20 e 29, respectivamente). No que respeita às formas moleculares, documenta-se para as M, 75 locais polimorficos e para as S, 88 locais polimorficos.

Tabela 5 - Numero de locais polimórficos e polimorfismos de ADN por região e forma molecular, nas zonas codificantes e não codificante do gene GNBPB3.

n° amostras	Locais polimorficos					Polimorfismos ADN												
	Exão 1	Exão 2	Intrão	3'	total	Exão 1		Exão 2		Intrão		Exão 1 + Exão 2						
	P	P	P	P	P	h	h	h	h	h	h	h	h					
Angola	6	3	17	4	1	20	2	0,008	0,006	5	0,010	0,034	2	0,010	0,025	5	0,010	0,007
Gabão	9	7	37	1	3	44	7	0,009	0,012	9	0,014	0,007	2	0,013	0,005	9	0,013	0,014
Ghana-A	7	5	38	3	6	43	5	0,009	0,009	7	0,016	0,020	3	0,015	0,017	7	0,015	0,015
Ghana-O	12	8	50	7	2	58	8	0,011	0,012	12	0,017	0,026	6	0,016	0,033	12	0,016	0,016
Guiné-B	19	8	68	2	0	76	8	0,008	0,011	19	0,015	0,004	3	0,014	0,008	19	0,014	0,018
Mozambique	10	4	25	4	1	29	3	0,007	0,006	7	0,011	0,023	3	0,011	0,020	7	0,011	0,009
Tanzânia	7	5	35	4	5	40	4	0,007	0,009	6	0,016	0,022	3	0,015	0,023	6	0,015	0,014
Formas M	24	8	67	5	3	75	9	0,010	0,010	23	0,015	0,023	4	0,015	0,019	23	0,015	0,017
Formas S	46	14	74	4	8	88	15	0,008	0,015	39	0,015	0,010	5	0,014	0,015	39	0,014	0,017
P- polimorfismos: h- haplotipos; π- diversidade nucleotídica; W- diversidade genética baseada em locais polimorficos																		

P- polimorfismos; h- haplotipos; π - diversidade nucleotídica; W- diversidade genética baseada em locais polimorficos

6.2 Testes de neutralidade

Testes da distribuição da frequência alélica nos locais polimórficos são frequentemente utilizados para calcular Tajima D. No entanto, não se obteve valores significativos de D. Sendo assim, não se pode determinar que o gene tenha experimentado recente actuação de selecção direccionada (ou purificadora) (anexo 1).

O teste Fu e Li não foi também significativo para todas as amostras ($p > 0,05$). (anexo 2)

Na análise do rácio K_a/K_s verifica-se que este rácio apresenta valores entre 0,024 e 0,049, muito abaixo de 1, ou seja, indicando a actuação da selecção purificadora. (tabela 6).

Tabela 6- Valores de K_a/K_s ao longo do gene GNBPB3

<i>Regiões</i>	K_a/K_s
ANGOLA -Gabão	0,03
ANGOLA- Ghana Accra	0,024
ANGOLA- Ghana Okyereko	0,031
ANGOLA- Guine Bissau	0,03
ANGOLA- Moçambique	0,034
ANGOLA- Tanzania	0,026
GABAO- Ghana Accra	0,029
GABAO- Ghana Okyereko	0,024
GABAO- Guine Bissau	0,028
GABAO- Moçamb	0,042
GABAO- Tanzania	0,023
GHANA A - Ghana Okyereko	0,033
GHANA A - Guine Bissau	0,033
GHANA A - Moçambique	0,049
GHANA A - Tanzania	0,032
GHANA O- Guine Bissau	0,039
GHANA O- Moçambique	0,044
GHANA O- Tanzania	0,028
GUINE B- Moçambique	0,05
GUINE B- Tanzania	0,036
MOÇAMB – Tanzânia	0,046
Formas M - Formas S	0,028
Ghana M/Ghana S	0,033
Guine M/Guine S	0,036
Ghana M/Guine M	0,044
Ghana S/Guine S	0,035
S ocidental/s oriental	0,029
S ocidental/m ocidental	0,027

A- Accra; O- Okyereko; B- Bissau
MOÇAMB- Moçambique

Como ao longo do gene podem existir impressões de uma pressão selectiva diferente do que a analisada no seu conjunto, foi efectuado um teste que analisa cada nucleótido em intervalos de 50 nucleotídeos, onde se verifica que no mesmo intervalo - de 1 a 75, entre as populações Ghana-Okyereko/Tanzânia; Angola/Ghana-Okyereko; Gabão/Moçambique; Ghana-Okyereko/Moçambique; Ghana-Okyereko/Guine; Moçambique/Tanzânia; Ghana-Accra/ Ghana/Okyereko e Gabão/Tanzânia, os valores de Ka/ Ks são superiores a 1 (figuras 4 a 10).

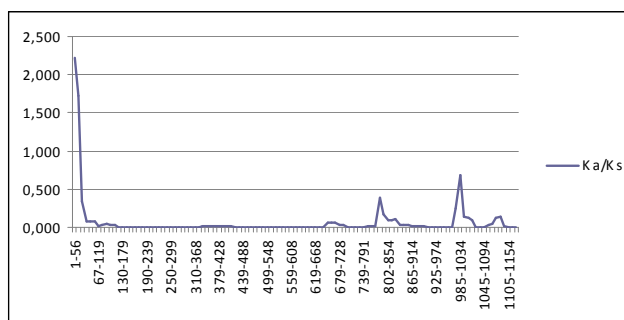


Figura 4 -Valores de Ka/Ks ao longo do gene GNBPB3 das formas M *versus* formas S em todas as populações

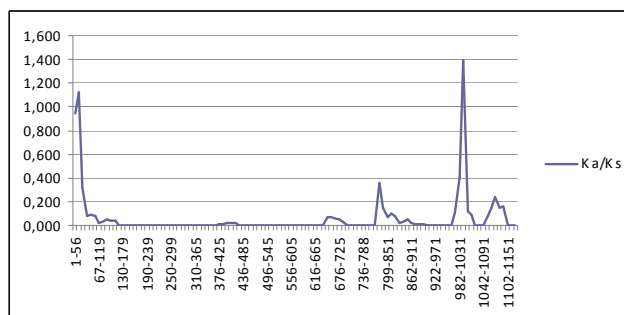


Figura 5 -Valores de Ka/Ks ao longo do gene GNBPB3 das formas M *versus* formas S no Ocidente

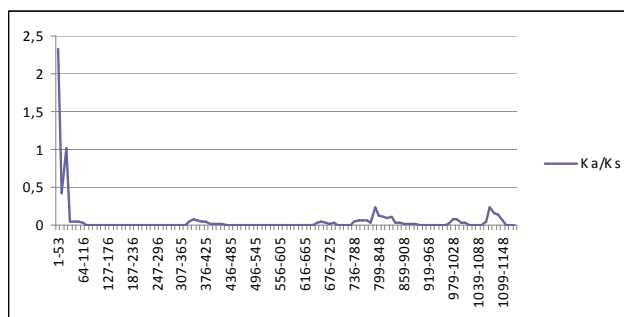


Figura6 - Valores de Ka/Ks ao longo do gene GNBPB3 das formas S Ocidente *versus* Oriente

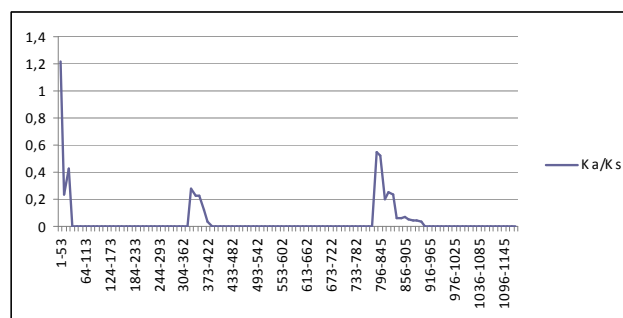


Figura7- Valores de Ka/Ks ao longo do gene GNBPB3
Gabão *versus* Moçambique

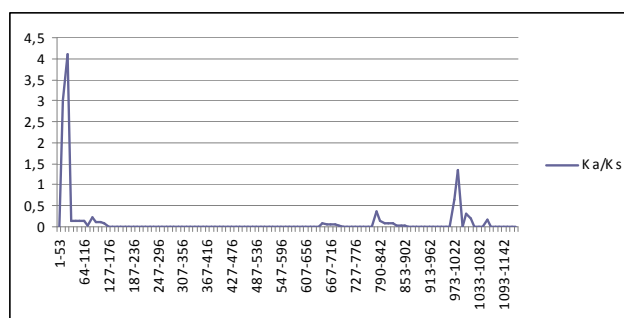


Figura 8 - Valores de Ka/Ks ao longo do gene GNBPB3
Ghana-Okyereko *versus* Guiné-Bissau

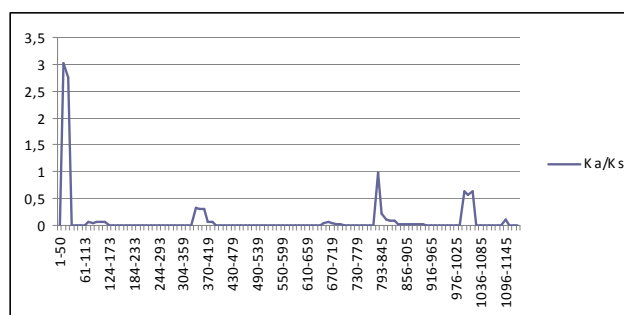


Figura9 - Valores de Ka/Ks ao longo do gene GNBPB3
Ghana-Okyereko *versus* Moçambique

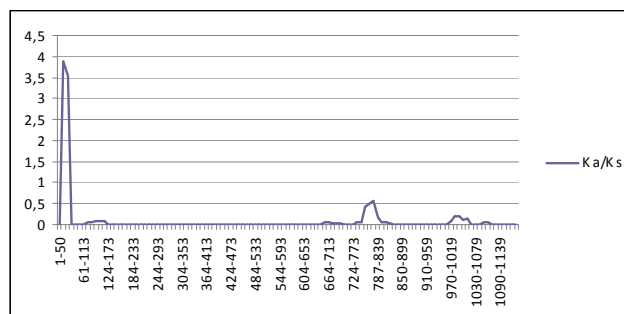


Figura10 - Valores de Ka/Ks ao longo do gene GNBPB3
Ghana-Okyereko *versus* Tanzânia

Através do programa informático (para análise de sequências peptídicas) SMARTgenome (<http://www.expasy.org>) confirma-se que a proteína em questão é uma glicosil hidrolase, que pertence à família 16 (figura 11). O intervalo em que os picos Ka/Ks são superiores a um, corresponde aos domínios sinal péptido e desordem intrínseca, da proteína (tabela 7).

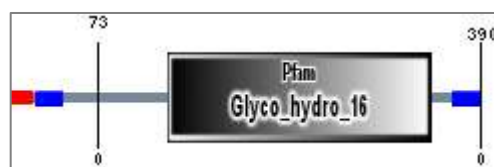


Figura 11 -Esquema da sequência peptídica GNBPB3.
(<http://smart.embl-heidelberg.de/>)

Tabela 7- Domínios da proteína GNBPB3

<i>Nome</i>	<i>Início</i>	<i>Fim</i>
Sinal péptido	1	19
Desordem intrínseca	20	43
Glicosil Hidrolase 16	130	349
Desordem intrínseca	367	390

Uma segunda análise foi efectuada com o teste de Tajima D em intervalos de 75 em 75 pares de bases. Neste teste há intervalos onde se registam valores inferiores a zero para um grau de significância $p < 0,05$, conferindo os valores obtidos anteriormente de estarmos perante a actuação da selecção purificadora (tabela 8).

Tabela 8- Valores do teste de Tajima D, em intervalos de 75 nucleótidos.

<i>Amostra</i>	<i>Intervalo (s) de significância</i>	<i>Valor de significância</i>	<i>D</i>
Total	[607-807]	#	-
	[1009-1158]	#	
Angola	[176-300];	#	+
	[751-875]	#	
Gabão	----	----	----
Ghana-Accra	[527-626]	#	+
Ghana-Okyereko	[352-451]	*	-
	[377-476]	*	
Guiné-Bissau	[177-276]	#	-
	[227-326]	#	
	[627-720]	*	
	[852-951]	*	
Moçambique	[251-352]	#	+
Tanzânia	----	----	----
Formas M	[627-720]		-
Formas S	[1007-1131]	#	-
	[1057-1156]	*	
Formas S Ocidente	[1006-1105]	#	-
	[1031-1130]	#	
	[1056-1155]	**	
Formas S Oriente	----	----	----

p <0,10 * p<0,05 ** p<0,01 ---- sem intervalos de significância + D>0 -D<0

O teste de McDonald e Kreitman não foi significativo para todas as amostras em estudo, pela ausência de substituições fixas (anexo 3).

6.3 Diferenciação genética e diversidade molecular

A caracterização e comparação da diferenciação genética nas diferentes zonas geográficas em estudo foram feitas utilizando o teste estatístico F_{st} entre todos os pares de regiões, separadas no caso da Guiné-Bissau e Ghana- Okyereko de acordo com a forma molecular (tabela 10) e sem separação da forma molecular (tabela 11).

Nem todos os valores encontrados foram significativos ($p < 0,05$), sendo os menores valores de F_{st} entre Ghana-Okyereko e Guiné-Bissau (0,054), Ghana-Accra e Moçambique (0,096) e entre Gabão-Dienga e Tanzânia (0,096) (tabela 11) entre Ghana_Okyereko M e Guiné-Bissau S (0,066) (tabela 10); Por outro lado o maior valor de F_{st} regista-se entre Moçambique e Guiné-Bissau_S (0,377) e entre Guiné-Bissau_S e Angola M (0,334). Verifica-se ainda que Moçambique apresenta diferenciação em relação a todas as regiões, excepto da Tanzânia (geograficamente menos afastado), assim como, o Gabão e Angola em que todas as comparações apresentam valores superiores a 0,05 (tabela 10) o que segundo Wright (1978), (tabela 9) corresponde a uma moderada diferenciação entre as amostras dos respectivos locais e em muitos dos casos os valores são mesmo superiores a 0,25. Efectuou-se ainda o teste entre as formas moleculares M e S onde se observa um F_{st} de 0,089, com significância ($p = 0$), correspondendo a moderada diferenciação, entre as variáveis estudadas (tabela 13).

Tabela 9 - Intervalos de diferenciação genética

<i>Intervalo</i>	<i>Interpretação (Wright, 1978)</i>
0-0,05	Pequena diferenciação
0,05-0,15	Moderada diferenciação
0,15-0,25	Bastante diferenciação
>0,25	Muita diferenciação

Tabela 10- Diferenciação genética interpopulacional (divisão da amostra de acordo com a região e forma molecular).

	<i>Moçambique</i>	<i>Tanzânia</i>	<i>Gabao_Dienga_S</i>	<i>Ghana_Accra_S</i>	<i>Ghana_Okyereko_S</i>	<i>Guine_Bissau_S</i>	<i>Guine_Bissau_M</i>	<i>Ghana_Okyereko_M</i>	<i>Angola_M</i>
Moçambique	0,000								
Tanzânia	-0,016	0,000							
Gabao_Dienga_S	0,137	0,096	0,000						
Ghana_Accra_S	0,096	0,038	0,056	0,000					
Ghana_Okyereko_S	0,168	0,110	0,107	0,005	0,000				
Guine_Bissau_S	0,377	0,284	0,282	0,156	0,086	0,000			
Guine_Bissau_M	0,300	0,227	0,226	0,128	0,125	0,021	0,000		
Ghana_Okyereko_M	0,201	0,121	0,118	0,017	0,021	0,066	-0,001	0,000	
Angola_M	0,308	0,257	0,257	0,137	0,207	0,334	0,246	0,131	0,000

A vermelho encontram-se os valores com grau de significância de $p < 0,05$.

Tabela 11- Diferenciação genética interpopulacional (divisão da amostra de acordo com a região).

	<i>Angola</i>	<i>Gabão</i>	<i>Ghana-Accra</i>	<i>Ghana-Okyereko</i>	<i>Guiné-Bissau</i>	<i>Moçambique</i>	<i>Tanzânia</i>
Angola	0,000						
Gabão	0,258	0,000					
Ghana-Accra	0,137	0,056	0,000				
Ghana-Okyereko	0,141	0,098	0,006	0,000			
Guiné-Bissau	0,265	0,246	0,145	0,054	0,000		
Moçambique	0,308	0,138	0,096	0,160	0,320	0,000	
Tanzânia	0,258	0,096	0,038	0,108	0,253	-0,016	0,000

A vermelho encontram-se os valores com significância de $p < 0,05$.

Tabela 12- Diferenciação genética interpopulacional formas M *versus* formas S no Ocidente

	<i>Angola_M</i>	<i>Ghana_Okyereko_M</i>	<i>Guine_Bissau_M</i>	<i>Gabao_Dienga_S</i>	<i>Ghana_Accra_S</i>	<i>Ghana_Okyereko_S</i>	<i>Guine_Bissau_S</i>
Angola_M	0,000						
Ghana_Okyereko_M	0,131	0,000					
Guine_Bissau_M	0,246	-0,001	0,000				
Gabão_Dienga_S	0,257	0,118	0,226	0,000			
Ghana_Accra_S	0,137	0,017	0,128	0,056	0,000		
Ghana_Okyereko_S	0,207	0,021	0,125	0,107	0,005	0,000	
Guine_Bissau_S	0,334	0,066	0,021	0,282	0,156	0,086	0,000

A vermelho encontram-se os valores com significância de $p < 0,05$.

Tabela 13- Diferenciação genética interpopulacional

	<i>Formas M</i>	<i>Formas S</i>
Formas M	0,000	
Formas S	0,089	0,000

A vermelho encontram-se os valores com significância de $p < 0,05$.

O teste da AMOVA foi efectuado de acordo com a seguinte hierarquia: Em primeiro lugar as formas moleculares (referido na tabela como entre grupos) seguidas das diferentes populações (referido na tabela como entre grupos dentro das populações) e por fim entre as amostras dentro de cada população (referido na tabela como dentro das populações). Verifica-se que os valores mais elevados dizem respeito ao segundo grau da hierarquia (diferentes localizações geográficas) e os menos entre as formas moleculares (tabela 14).

Tabela 14-AMOVA comparando todas as populações e as diferentes regiões e formas moleculares.

<i>AMOVA</i>	<i>M vs S</i> <i>localização</i> <i>geográficas</i>	<i>M vs S</i> <i>localização</i> <i>geográfica só</i> <i>ocidente</i>	<i>S Ocidente vs S Oriente</i> <i>localizações</i> <i>geográficas</i>	<i>M vs S</i> <i>Ghana-Okyereko e</i> <i>Guine</i>
Entre grupos	5,65	1,95	9,19	3,07
Entre grupos dentro das populações	13,03	12,51	8,40	3,19
Dentro das populações	81,32	85,55	82,41	93,74
FSC	0,13810*	0,12756*	0,09194*	0,03290*
FST	0,18676*	0,14453*	0,17591*	0,06261
FCT	0,05646	0,01946	0,09248	0,03072

* Valor de p < 0,05

6.4 Análise filogenética

A primeira rede de haplotipos (figura 12) está construída de acordo com os haplotipos referentes às amostras das sete populações em estudo (correspondendo cada cor a uma população). Observa-se uma grande variedade de haplotipos (63), distribuídos pelas populações, onde apenas cinco deles são partilhados por mais do que uma população e somente dois destes são partilhados por populações diferentes e um por mais de duas amostras, como se pode observar na figura pelos círculos partilhados representados com mais do que uma cor. Por exemplo, o círculo que na figura se denomina MZ27, corresponde a um haplotipo partilhado por três amostras provenientes de Moçambique e uma amostra da Tanzânia, enquanto que o círculo denominado TZ209A, corresponde a um haplotipo partilhado por duas amostras diferentes da Tanzânia.

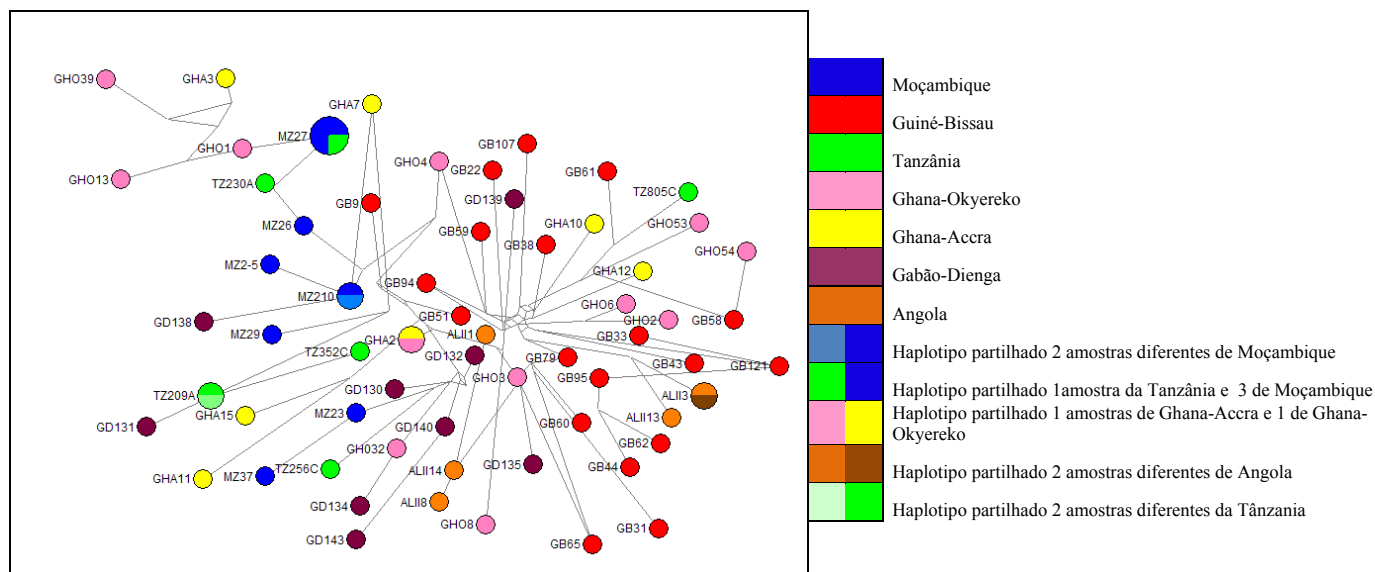


Figura 12 - Rede de haplotipos referente às diferentes amostras das sete populações em estudo.

Cada cor corresponde a uma população.

De seguida, construiu-se uma rede de haplotipos, tendo em consideração a variável forma molecular, onde se documenta uma separação haplotípica entre as duas formas moleculares (figura 13) contrariando a tendência dos resultados obtidos nos outros testes realizados.

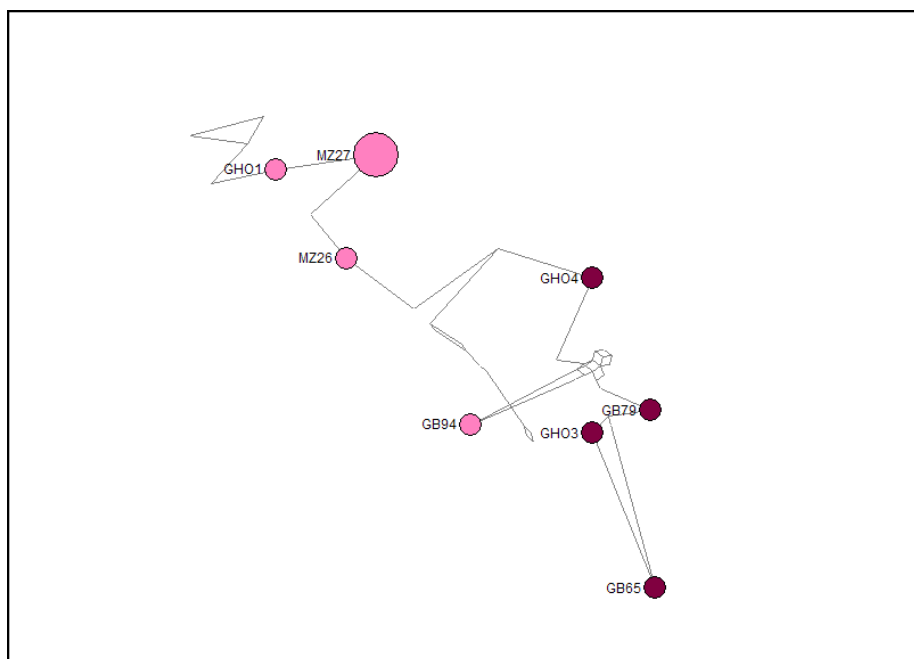


Figura13- Torso da rede de haplotipos referente às formas moleculares

O rosa claro corresponde à forma molecular S e o rosa escuro à forma molecular M.

Por fim, construiu-se uma rede de haplotipos (figura 14) tendo em consideração as variáveis forma molecular e localização geográfica (Oriente *versus* Ocidente), onde se verifica uma separação não só entre as duas formas como também entre as zonas geográficas. Moçambique e a Tanzânia (representadas a verde) são as únicas regiões, em estudo, da África Oriental e forma molecular S, que tendencialmente estão mais próximas das formas S restantes (representadas a rosa claro). Nas figuras 13 e 14 apenas está esquematizado o torso central da rede. (esquemas originais anexos 4 e 5, respectivamente).

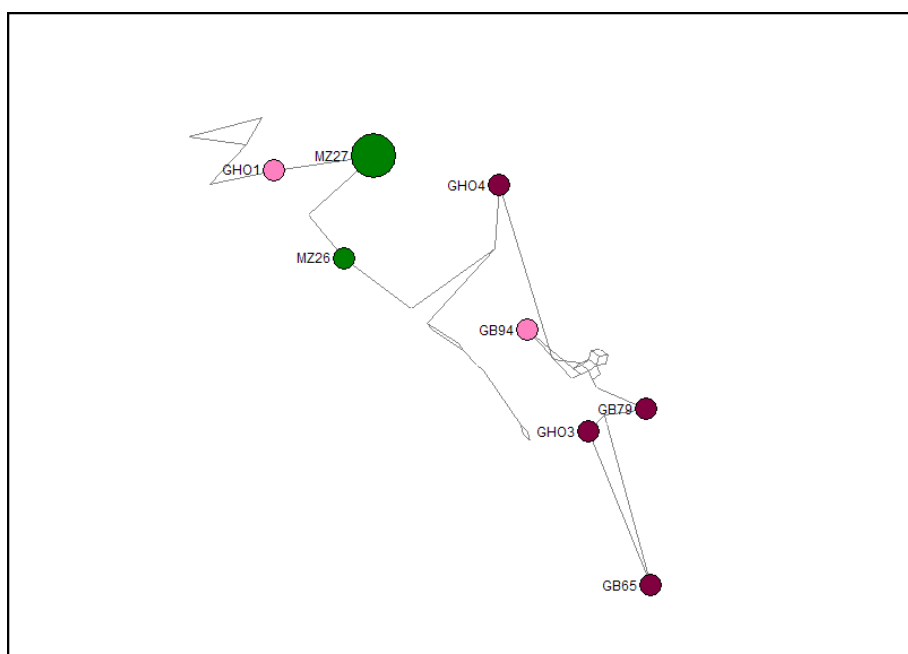


Figura14- Torso da rede de haplotipos referente às formas moleculares e Ocidente *versus* Oriente
O rosa claro corresponde à forma molecular S da zona Ocidental; a verde a forma molecular S da Zona Oriental e o rosa escuro à forma molecular M.

7 Discussão

A prevalência alarmante da malária, associada ao facto de não existir uma vacina efectiva (Richie & Saul, 2002), da resistência do *P. falciparum* aos anti-maláricos e da resistência do *A. gambiae* aos insecticidas (Olliaro, 2001) torna urgente novas medidas de controlo da doença, despertando especial interesse para as relações parasita-mosquito, especialmente no que se refere à resposta imunológica do vector, que transversalmente com vários estudos neste contexto, poderá ser responsável por grande parte das perdas parasitárias (Osta *et al.*, 2004).

Compreender a dinâmica hospedeiro-patogénio, torna-se essencial para entender a evolução do sistema imunológico, uma vez que este desempenha um papel fundamental na mediação destas relações, remetendo para uma série de interacções ao longo da história evolutiva dos dois agentes, onde qualquer avanço evolutivo no patogénio vai causar pressão selectiva no respectivo hospedeiro, pois, este tem necessidade de restabelecer a sua capacidade de resposta, experimentando uma adaptação recíproca (Lazzaro, 2008). Desta dinâmica pode resultar uma forte assinatura de adaptação evolutiva nos genes envolvidos na resposta imunológica, uma vez que estes poderão ser alvo de selecção direccional ou balanceadora mais do que os genes que não estão envolvidos nesta resposta (Sackton *et al.*, 2003). Um exemplo são os genes envolvidos no sistema imunológico dos vertebrados e da *Drosophila* que exibem altas taxas de evolução nos aminoácidos, quando comparados com genes que não fazem parte do sistema imunológico (Jiggins & Hurst, 2003), indicando adaptação molecular.

No entanto, nem todos os genes evoluem do mesmo modo, podendo servir como base para inferir até que ponto as interacções de diferentes tipos de parasitas com os vários componentes do sistema imunológico interferem na história evolutiva dos diversos genes reguladores da resposta imunológica. Então, o tipo de selecção e o modo como estes genes evoluem, delineia a capacidade do mosquito para eliminar os microorganismos invasores e consequentemente de sobreviver. Desvios à neutralidade podem indicar que a população está sob pressão selectiva.

A grande maioria das combinações *Plasmodium*-vector é incompatível e são muitos os factores que contribuem para a capacidade inata de algumas espécies de anofelinos suportarem o crescimento e o desenvolvimento de determinadas espécies de *Plasmodium*. (Warburg & Miller, 1991). As incompatibilidades referidas podem advir de barreiras biológicas ou fisiológicas do mosquito, ou pela activação do seu sistema imunológico inato, resultando, por exemplo, na morte de determinados estadios durante o desenvolvimento do parasita. O *P. falciparum* sofre um grande

número de perdas durante o seu desenvolvimento no interior do *A. gambiae* (Warburg & Miller, 1991). Segundo a literatura, um dos factores que para isto contribui é a existência de determinados genes efectores na resposta imunológica do mosquito, responsáveis pela produção de proteínas envolvidas na eliminação ou mesmo no bloqueio do ciclo de vida do parasita (Moreira *et al.*, 2002).

O gene em estudo neste trabalho está envolvido na resposta imunitária do *A. gambiae*. Entender a sua evolução e o modo como a selecção natural actua é essencial para determinar até que ponto é responsável pela eliminação ou bloqueio do ciclo de vida do *P. falciparum*. A diversidade nucleotídica (π) obtida neste estudo foi relativamente alta para todas as populações - correspondendo o maior valor a Ghana-Okyereko ($\pi = 0,016$) - quando comparadas com o nível de diversidade nucleotídica observada no genoma humano ($\pi < 0,006$) (The International SNP Map Working Group, 2001). No entanto, são valores semelhantes aos obtidos anteriormente em estudos de genes envolvidos na resposta imunológica da *Drosophila* e de *A. gambiae* (Simard F. *et al.*, 2007; Morlais I. *et al.*, 2004; Obbard D.J. *et al.*, 2007; Andolfatto P., 2001) e inclusive para o gene da GNBPB3 num estudo efectuado por Cohuet *et al.* em 2008 ($\pi = 0,017$).

No que se refere às formas moleculares, valores de π são muito semelhantes documentando-se 0,015 para as formas M e 0,014 para as formas S (no nosso estudo). Quando comparado com o estudo realizado por Cohuet *et al.* (2008) à escala genómica, as formas M apresentam valores ligeiramente superiores enquanto que as S são muito semelhantes (M= 0,008 e S= 0,015). As diferenças entre os dois estudos podem ser consequência das diferentes localizações geográficas, até porque no nosso estudo as formas moleculares M estão apenas representadas por três países da África ocidental (Angola; Ghana-Okyereko e Guiné-Bissau), e a forma molecular S distribuída por seis zonas de África ocidental e oriental.

Relativamente à diferenciação de populações, o f_{st} mostra a maioria dos valores significativos $p < 0,05$ quando se comparam todas as populações, independentemente da forma molecular, verifica-se que todas as populações estão geneticamente separadas, observando-se que Angola, Moçambique e Guiné-Bissau são as populações geneticamente mais afastadas e que o maior valor de f_{st} é entre Moçambique e Guiné-Bissau (0,32) que são as zonas geograficamente mais afastadas. Provavelmente o aumento da distância entre as populações é factor determinante nas diferenças exercidas pela pressão selectiva, e consequentemente numa maior diferenciação entre as populações. A confirmar esta hipótese o menor valor de f_{st} é entre as populações que são geograficamente mais próximas: Ghana-Okyereko e Guiné-Bissau (0,054), Ghana-Accra e Guiné-Bissau (0,145). É interessante assinalar que quando se compara Ghana-Accra (forma S) com Guiné

Bissau (forma M e S), apesar de baixo, o maior valor de f_{st} corresponde à comparação Ghana-Accra com Guiné-Bissau S (0,156) em vez de Ghana-Accra com Guiné-Bissau M (0,128), o que pode indicar que a forma molecular contribui menos para a diferenciação molecular das amostras em estudo do que a sua localização geográfica. A consolidar esta afirmação, no estudo das formas moleculares M e S, na costa Ocidental do continente Africano, verifica-se que o menor valor de π (0,067) corresponde à comparação entre Ghana-Okyereko M e Guiné-Bissau S, ou seja, um valor muito baixo de diferenciação genética- entre duas regiões próximas e ao mesmo tempo as duas formas moleculares distintas. Neste mesmo teste observam-se valores altos de f_{st} entre populações distantes com a mesma forma molecular, como é o caso de Gabão-Dienga S e Guiné-Bissau S e Angola M e Guiné-Bissau M (0,282 e 0,246, respectivamente). Quando se efectua o estudo entre as duas formas moleculares (independentemente da localização) documenta-se um valor baixo de f_{st} de 0,08, correspondendo a diferenciação moderada entre as populações. Este facto pode mais uma vez dever-se à grande variedade de regiões em estudo.

Comparando Ghana-Okyereko M e S, Guiné-Bissau M e S e Ghana-Accra com Ghana-Okyereko, (que correspondem à mesma área geográfica) obtém-se os valores de f_{st} mais baixos (0,021, 0,021 e 0,006, respectivamente), inferiores a 0,05, que corresponde a pequena diferenciação, no entanto, não é possível validar uma vez que para estes casos o valor de p não é significativo.

Os resultados obtidos na AMOVA corroboram as afirmações anteriores uma vez que a variação molecular é maior quando se compara as regiões (Entre grupos dentro das populações) do que quando se compara as formas moleculares M e S (entre grupos).

Relativamente aos resultados dos testes de neutralidade, Tajima D e Fu & Li não foram significantes para desvios à neutralidade, bem como o teste de McDonald-Kreitman que demonstrou um défice de substituições fixas não-sinónimas, para todas as populações.

Os valores obtidos para o rácio Ka/Ks foram inferiores a 1, indicando que o gene GNBPB3 está sob o efeito de selecção purificadora à semelhança do que aconteceu em estudos anteriores para este mesmo gene, por exemplo no estudo efectuado por Cohuet *et al.*, em 2008 onde o valor Ka/Ks entre as formas moleculares M e S é o mesmo obtido neste estudo (0,028) e em estudos efectuados por Bulmer & Crozier (2006) e Little *et al.* (2004) onde verificaram que os genes que codificam as GGBP e as PGRP's de *Daphnia* e *Drosophila* estão sob a acção da selecção purificadora, como seria de esperar uma vez que os peptidoglicanos, como componentes da membrana celular das bactérias, não podem alterar a sua composição para se tornarem indetectáveis para o hospedeiro,

isto deve-se da natureza conservadora dos padrões moleculares associados aos patógenos (PAMPs), como por exemplo dos lipopolissacarídeos e peptidoglicanos, que estão restringidos evolutivamente, pois não podendo alterar as suas propriedades bioquímicas (Loker *et al.* 2004). Este tipo de selecção contribui para a eliminação de variações nos aminoácidos, isto é, quando existe uma diminuição da diversidade genética, à medida que se verifica uma estabilização da população, por eliminação de alelos ou mutações raras, estamos sob a acção da selecção negativa ou purificadora. Por outro lado, quando a selecção natural favorece um único alelo e consequentemente a frequência alélica é alterada nesta direcção, estamos perante selecção positiva ou direcciona (Koonin & Galperin, 2003).

Através do estudo efectuado, ao longo do gene, dos locais polimórficos, verificou-se um maior número de polimorfismos nas regiões codificantes do que nas zonas não codificantes. Os polimorfismos de um único nucleótido (*Single nucleotide polymorphisms* – SNP's) são considerados a forma mais comum de variação genética nos vertebrados e invertebrados (Cohuet *et al.*, 2008) sendo por isso um dos veículos escolhido para o estudo genético das populações e para identificar se os genes estão sob a pressão de pressão selectiva. Nas regiões codificantes do gene, os SNP's podem comportar-se de duas formas: substituições sinónimas ou silenciosas, ou seja, quando ocorre a substituição de um nucleótido que não vai alterar a estrutura do aminoácido e consequentemente da proteína ou substituições não-sinónimas, quando a substituição ocorre de modo a alterar o aminoácido e consequentemente a proteína. (Cohuet *et al.*, 2008). Neste estudo não se observaram substituições não-sinónimas na região codificante. Por outro lado, verificou-se a presença de substituições silenciosas, ao longo do gene. Este facto apoia a afirmação de que o gene da GNBPB3 na sua totalidade encontra-se sob a acção de selecção purificadora, uma vez que perante a acção da selecção positiva é esperado um aumento do número de substituições não-sinónimas (McDonald & Kreitman, 1991), comparativamente às sinónimas aliado à baixa diversidade da região codificante (Simard *et al.*, 2007).

No estudo efectuado por Bulmer M.S. & Crozier R.H. (2006), estes verificaram a existência de selecção positiva nos genes GGBP1 e GGBP2 da térmita *Nasutitermes comatus*, que sugerem estar relacionado com os locais de deleção e substituições de aminoácidos relacionados com a activação da proteína. Tendo em consideração que os patógenos podem interferir com a habilidade das proteínas de reconhecimento transmitirem sinal, consequentemente existe a possibilidade de ocorrer evolução adaptativa (Begun & Whitley 2000). Segundo o modelo “arms race” da co-evolução (Dawkins R. & Krebs J.R., 1979), seria de esperar que, uma vez que os patógenos exercem uma pressão selectiva no sistema imunológico do hospedeiro, existam sinais de pressão selectiva

direcciona. O modelo “arms race” assenta na seguinte hipótese: se os patógenos estão constantemente a encontrar mecanismos para se evadirem da resposta imunológica do hospedeiro e por seu lado os hospedeiros têm necessidade de adoptarem mecanismos mais eficientes no controlo das infecções a que estão sujeitos, esta dinâmica promoveria a rápida evolução dos genes envolvidos nas interacções parasita-hospedeiro, com um consequente aumento da frequência de alelos, que do ponto de vista da selecção se tornam vantajosos nestas interacções. Estudos descrevem em *Drosophila* a presença de determinados genes, envolvidos na resposta imunológica, que apresentam sinais de selecção direcciona ou positiva (Lazzaro B.P. & Clark A.G., 2003; Little T.J. & Cobbe N., 2005; Jiggins F.M & Kim K.W, 2007). Contudo, mais recente Cohuet e colaboradores (2008), não detectaram padrão de selecção direcciona à excepção do gene TEP 1, em que verificou valores positivos para o teste de Tajima D, sugerindo a conservação de alelos divergentes.

No nosso estudo, quando comparamos o rácio Ka/Ks ao longo do gene, verificou-se a existência de uma pressão selectiva diferente – selecção positiva – do que quando analisamos o gene na sua totalidade – selecção purificadora – no intervalo de [1 a 75 nucleotidos da região codificante] para as populações Ghana-Okyeroko/Tanzânia; Angola/Ghana-Okyeroko; Gabão/Moçambique; Ghana-Okyeroko/Moçambique; Ghana-Okyeroko/Guine; Moçambique/Tanzânia; Ghana-Accra/Ghana-Okyeroko e Gabão/Tanzânia. Verifica-se então que estamos perante a acção da selecção positiva e que este intervalo corresponde aos domínios de sinal péptideo e de desordem intrínseca da proteína. A desordem intrínseca refere-se a uma região de uma proteína onde existe como um conjunto estrutural, tanto ao nível da sua conformação secundária como terciária, onde não consegue alcançar a sua estrutura tridimensional (Dunker, A.K. et al., 2001). As regiões de desordem intrínseca estão envolvidas na regulação, sinalização e reconhecimento, e parecem ter um papel importante na regulação postraducional como a acetilação, hidroxilação, metilação e fosforilação. Este facto sugere que as regiões do gene envolvidas no reconhecimento estão sobre pressão da selecção purificadora, devido aos constrangimentos impostos pelas PAMPs, enquanto que as regiões associadas à sua regulação e secreção poderão estar sujeitas a pressões selectivas mais em consonância com o modelo evolutivo “arms race”.

8 Conclusão

Neste estudo observou-se que todas as populações estão geneticamente separadas e que a distância entre as populações é provavelmente um factor determinante nas diferenças exercidas pela pressão selectiva, resultando assim, numa maior diferenciação entre as populações. Ao mesmo tempo, verificou-se que a forma molecular contribui menos para a diferenciação molecular das amostras em estudo, do que a sua localização geográfica.

Os resultados obtidos, estão de acordo com estudos anteriores, no contexto de genes relacionados com o sistema imunológico inato tanto de *A. gambiae* (Simard F. *et al.*, 2007, Little & Cobbe, 2005) como de *Drosophila* (Jiggins & Hurst, 2003), no que se refere ao tipo de selecção a actuar. Verificou-se que o gene GNBPB3 está sob a acção da selecção purificadora. No entanto, identificaram-se, em algumas das populações, sinais da presença de selecção positiva na região 5' do gene, em estudo. Este tipo de acontecimento já havia sido reportado em estudos anteriores (Bulmer M.S. & Crozier R.H., 2006) para genes das GGBP em térmitas (GNBP1 e GNB2) e segundo os autores, em muito se relacionava com a variedade de microorganismos nos diferentes habitats que experimentavam.

Neste estudo, e apesar de deparamos com uma grande variedade de regiões, com características ecológicas distintas, os resultados obtidos remetem como já referido, para a acção de selecção purificadora em todas as populações.

9 Referências bibliográficas

- Alavi, Y., Arai, M., Mendoza, J., Tufet-Bayona, M., Sinha, R., Fowler, K., Billker, O., Franke-Fayard, B., Janse, C. J., Waters, A. & Sinden, R. E., 2003. The dynamics of interactions between *Plasmodium* and the mosquito, a study of the infectivity of *Plasmodium berghei* and *Plasmodium gallinaceum*, and their transmission by *Anopheles stephensi*, *Anopheles gambiae* and *Aedes aegypti* International Journal for Parasitology, **33** (9): 933-943.

- Andolfatto P, 2001. Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. Mol Biol Evol, **18**(3): 279-290.

- Barillas-Mury, C., Charlesworth, A., Gross, I., Richman, A., Hoffman, J. A. & Kafatos, F. C., 1996. Immune factor Gambif1, a new rel family member from the human malaria vector, *Anopheles gambiae*. EMBO J. **15**: 4961–4701.

- Baxter RH, Chang CI, Chelliah Y, Blandin S, Levashina EA, Deisenhofer J., 2007. Structural basis for conserved complement factor-like function in the antimalarial protein TEP1. Proc. Natl. Acad. Sci. USA, **104**:11615-11620.

- Begun, D. J., & P. Whitley, 2000. Adaptive evolution of Relish, a *Drosophila* NF- κ B/I κ B protein. Genetics **15**(4): 1231–1238.

- Beier, J.C., Straif, S. C., Mbogo, C. N., Toure, A. M., Walker, E. D., Kaufman, M. & Toure, Y. T., 1998. Midgut bacteria in *Anopheles gambiae* and *An. funestus* (Diptera: Culicidae) from Kenya and Mali. J. Med. Entomol. **35**:222-226

- Blandin S, Shiao SH, Moita LF, Janse CJ, Waters AP, Kafatos FC, Levashina EA, 2004: Complement-like protein TEP1 is a determinant of vectorial capacity in the malaria vector *Anopheles gambiae*. Cell, **116**(5):661-670.

- Blandin, S., Moita, L. F., Kocher, T., Wilm, M., Kafatos, F. C. & Levashina, E. A., 2002. Reverse genetics in the mosquito *Anopheles gambiae*, targeted disruption of the Defensin gene. EMBO Rep. **3**: 852-856.

- Brey P.T., Ahmed A., Lee W-J., Ashida M., Lehane M.J., 1995. Tyrosinase-type prophenoloxidase distribution in the alimentary canal of strains of *Anopheles gambiae* refractory and susceptible to *Plasmodium* infection. *Exp. Parasitol.*, **80**(4):654-664.

- Bulmer , Mark S. & Crozier , Ross H., 2006. Variation in Positive Selection in Termite GNBPs and Relish. *Mol. Biol. Evol.* **23**(2):317–326.

- Christophides, G.K., Zdobnov, E., Barillas-Mury, C., Birney, E., Blandin, S., Blass, C., Brey, P.T., Collins, F.H., Danielli, A., Dimopoulos, G., Hetru, C., Hoa, N.T., Hoffmann, J.A., Kanzok, S.M., Letunic, I., Levashina, E.A., Loukeris, T.G., Lycett, G., Meister, S., Michel, K., Moita, L.F., Muller, H.M., Osta, M.A., Paskewitz, S.M., Reichhart, J.M., Rzhetsky, A., Troxler, L., Vernick, K.D., Vlachou, D., Volz, J., von Mering, C., Xu, J., Zheng, L., Bork, P., Kafatos, F.C., 2002. Immunity-related genes and gene families in *Anopheles gambiae*. *Science* **298**:159-165.

- Cohuet, Anna; Krishnakumar, Sujatha, Simard, Frédéric; Morlais, Isabelle, 2008. SNP discovery and molecular evolution in *Anopheles gambiae*, with special emphasis on innate immune system. *BMC Genomics* **9**: 227.

- Collins F.H., Finnerty V. & Petrarca V., 1988. Ribosomal DNA-probes differentiate five cryptic species in the *Anopheles gambiae* complex. *Parassitologia* **30**: 231-240.

- Collins, F.H. & Besansky, N.J., 1994. Vector biology and the control of malaria in Africa. *Science*, **264**: 1874–1875.

- Collins, F.H. & Paskewitz, S.M., 1995. Malaria: current and future prospects for control. *Annu. Rev. Entomol.*, **40**: 195– 219.

- Collins, W.E., Walduck, A., Sullivan, J.S., Andrews, K., Stowers, A., Morris, C.L., Jennings, V., Yang, C., Kendall, J., Lin, Q., Martin, L.B. Diggs, C. & Saul, A., 2000. Efficacy of vaccines containing rhoptry-associated proteins RAP1 and RAP2 of *Plasmodium falciparum* in *Saimiri boliviensis* monkeys. *Am J Trop Med Hyg* **62**: 466-479

- Coluzzi M., Di Deco, M.A., Petrarca, V., 1985. Chromosomal inversion intergradation and incipient speciation in *Anopheles gambiae*. *Boll Zool*, **52**:45-63.

- Dawkins, R., Krebs, J.R., 1979: Arms races between and within species. *Proc R Soc Lond B Biol Sci*, 205(1161):489-511.

- Diabaté, Abdoulaye, Dabiré, Roch K., Hultmark & Kyle D., 1993. Immune reactions in *Drosophila* and other insects: a model for innate immunity. *Trends Genetics* **9**:178–183.

- Dimopoulos, G., Kumar, S., Christophides, G. K., Cantera, R., Charles, B., Han, Y. S., Meister, S., Kafatos, F. C. & Barillas-Mury, C., 2003. The role of reactive oxygen species on *Plasmodium* melanotic encapsulation in *Anopheles gambiae*. *Proc. Natl. Acad. Sci.*, **100** (24): 14139-14144.

- Drakeley, C.; Schellenberg, D.; Kihonda, J.; Sousa, C.A.; Arez, A.P.; Lopes, D.; Lines, J.; Mshinda, H.; Lengeler, C.; Armstrong, S.J.; Tanner, M.; Alonso, P., 2003. An estimation of the entomological inoculation rate for Ifakara: a semi-urban area in a region of intense malaria transmission in Tanzania. *Trop Med. Int. Health*. **8**:767–774.

- Dunker, A. Keith, Lawson, David J., Brown, Celeste J., Williams, Ryan M., Romero, Pedro, Oh, Jeong S., Oldfield, Christopher J., Campen, Andrew M., Ratliff Catherine M., Hipps, Kerry W., Ausio, Juan, Nissen, Mark S., Reeves, Raymond, Kang, ChulHee, Kissinger, Charles R., Bailey, Robert W., Griswold Michael D., Chiu, Wah, Garner, Ethan C. and Obradovic, Zoran, 2001. Intrinsically disordered protein. *Journal of Molecular Graphics and Modelling* **19**(1): 26-59.

- Dziarski, R., 2004. Peptidoglycan recognition proteins (PGRPs). *Mol. Immunol.*, **40**:877-886.

- Evans, Andrew G. & Wellems, Thomas, E., 2002. Coevolutionary Genetics of *Plasmodium* Malaria Parasites and Their Human Hosts. *Integ. and Comp. Biol.*, **42**:401–407

- Excoffier, L., Smouse, P. E., & Quattro, J. M., 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **131**: 479-491.

- Fanello C., Santolamazza F., della Torre A., 2002. Simultaneous identification of species and molecular forms of the *Anopheles gambiae* complex by PCR-RFLP. *Med Vet Entomol.* **16**: 461–464.

- Fu, Y.-X. & W.-H. Li, 1993. Statistical tests of neutrality of mutations. *Genetics* **133**: 693-709.

- Greenwood, M., Fidock, David A., Kyle, Dennis E., Kappe , Stefan H.I., Alonso , Pedro L., Collins , Frank H. & Duffy, Patrick E., 2008. Malaria: progress, perils, and prospects for eradication. *The Journal of Clinical Volume* **118** (4):1266-1276.

- Grossman, G.L., Rafferty, C.S., Clayton, J.R., Stevens, T.K., Mukabayire, O., Benedict, M.Q., 2001. Germline transformation of the malaria vector, *Anopheles gambiae*, with the piggyBac transposable element. *Insect Mol. Biol.* **10**: 597–604.

- Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* **41**:95-98.

- Hoffmann, J. A., Reichhart, J. M. & Hetru, C., 1996. Innate immunity in higher insects. *Curr. Opin. Immunol.*; **8**(1):8–13.

- Hoffmann, J.A. & Reichhart, J.-M., 2002. *Drosophila* innate immunity: An evolutionary perspective. *Nature Immunology* **3**:121-126.

- Holt, R.A., Subramanian, G.M., Halpern A., Sutton, G.G., Charlab, R, Nusskern, D.R., Wincker, P., Clark, A.G., Ribeiro, J.M., Wides, R., Salzberg, S.L., Loftus B, Yandell M, Majoros W.H., Rusch D.B., Lai Z, Kraft CL, Abril JF, Anthouard V, Arensburger P, Atkinson P.W., Baden H., de Berardinis V., Baldwin D., Benes V, Biedler J., Blass C., Bolanos R., Boscus, D., Barnstead, M., Cai, S., Center, A., Chaturverdi, K., Christophides, G.K., Chrystal, M.A., Clamp, M., Cravchik, A., Curwen, V., Dana, A., Delcher, A., Dew, I., Evans, C.A., Flanigan, M., Grundschober-Freimoser A., Friedli, L., Gu, Z., Guan, P., Guigo, R., Hillenmeyer, M.E., Hladun, S.L., Hogan, J.R., Hong, Y.S., Hoover, J., Jaillon, O., Ke, Z., Kodira, C., Kokoza, E., Koutsos, A., Letunic, I., evitsky, A., Liang, Y, Lin, JJ, Lobo, NF, Lopez, JR, Malek, JA, McIntosh, TC, Meister, S, Miller J, Mobarry, C., Mongin, E., Murphy, S.D., O'Brochta, D.A., Pfannkoch C., Qi R., Regie,r M.A., Remington, K., Shao,

- H., Sharakhova, M.V., Sitter, C.D., Shetty J., Smith, T.J., Strong, R., Sun J., Thomasova, D., Ton, L.Q., Topalis P., Tu Z., Unger M.F., Walenz B., Wang, A., Wang, J., Wang, M., Wang, X., Woodford, K.J., Wortman, J.R., Wu, M., Yao A, Zdobnov, E.M., Zhang H., Zhao, Q., Zhao S., Zhu, S.C., Zhimulev, I., Coluzzi, M., della Torre, A., Roth, C.W., Louis, C., Kalush, F., Mural, R.J., Myers, E.W., Adams M.D., Smith, H.O., Broder, S., Gardner, M.J., Fraser, C.M., Birney, E., Bork, P., Brey, P.T., Venter, J.C., Weissenbach, J., Kafatos, F.C., Collins, F.H., Hoffman S.L., 2002: The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science*, **298**(5591):129-149.
- Jiggins, Francis M. & Hurst, Gregory D. D., 2003. The Evolution of Parasite Recognition Genes in the Innate Immune System: Purifying Selection on *Drosophila melanogaster* Peptidoglycan Recognition Proteins. *Journal of Molecular Evolution*, **57**(5): 598-605.
 - Jiggins, F.M. & Kim, K.W., 2007: A screen for immunity genes evolving under positive selection in *Drosophila*. *J. Evol. Biol.*, **20**(3): 965-970.
 - Kim, Y. S., Ryu, J. H., Han, S. J., Choi, K. H., Nam, K. B., Jang, I. H., Lemaitre, B., Brey, P. T. & Lee, W. J., 2000. Gram-negative bacteria binding protein, a pattern recognition receptor for lipopolysaccharide and beta-1,3-glucan that mediates the signaling for the induction of innate immune genes in *Drosophila melanogaster* cells. *J. Biol. Chem.* **275**, 32721- 32727.
 - Kimura, M., 1968. Evolutionary Rate at the Molecular Level. *Nature* **217**:624-626.
 - Koonin, Eugene V. & Galperin, Michael Y., 2003. Sequence-Evolution-Function: Computational Approaches in Comparative Genomics. Kluwer Academic Publishers. Cap. 2.1.
 - Knell, A. J., 1991 – Malaria. A publication of the Tropical programme of the Wellcome Trust. Oxford University Press. pp. 19-24
 - Lambrechts, L., Halbert, J., Durand, P., Gouagna, Louis C. & Koella, Jacob C., 2005. Host genotype by parasite genotype interactions underlying the resistance of anopheline mosquitoes to *Plasmodium falciparum*, *Malar J*; **4**: 3.

- Langer, R. C., Hayward, R. E., Tsuboi, T., Tachibana, M., Torii, M. and Vinetz, J. M., 2000). Micronemal transport of *Plasmodium* ookinete chitinases to the electron-dense area of the apical complex for extracellular secretion. *Infect. Immun.* **68**: 6461-6465.
- Lanzaro, G.C. & Tripet, F., 2003. Gene flow among populations of *Anopheles gambiae*: A critical review. In: *Ecological Aspects for the Application of Genetically Modified Mosquitoes*. Takken, Wageningen Frontis Series, Kluwer Academic Press, Dordrecht., **2**:109-132.
- Lazzaro, B.P., Clark AG, 2003: Molecular population genetics of inducible antibacterial peptide genes in *Drosophila melanogaster*. *Mol Biol Evol*, **20**(6):914-923.
- Lazzaro, Brian P., Schlenke, Todd A., Evans, Jay D., Sackton, Timothy B., Hultmark, Dan & Clark, Andrew G., 2007. Dynamic evolution of the innate immune system in *Drosophila*, *Nature Genetics*, **39**(12): 1461-1468.
- Lazzaro, Brian P, 2008. Natural selection on the *Drosophila* antimicrobial system *Current opinion in Microbiology*, **11**:284-289.
- Lee, W. J., Lee, J. D., Kravchenko, V. V., Ulevitch, R. J. and Brey, P. T., 1996. Purification and molecular cloning of an inducible gram-negative bacteria-binding protein from the silkworm, *Bombyx mori*. *Proc. Natl. Acad. Sci.* **93**(15): 7888–7893.
- Levashina, E. A., Moita, L. F., Blandin, S., Vriend, G., Lagueux, M. and Kafatos, F. C., 2001. Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, *Anopheles gambiae*. *Cell* **104**: 709-718.
- Little, T.J. & Cobbe, N., 2005: The evolution of immune-related genes from disease carrying mosquitoes: diversity in a peptidoglycan- and a thioester-recognizing protein. *Insect Mol Biol*, **14**(6):599-605.
- Little, T. J., Colbourne, J. K. & Crease, T. J., 2004. Molecular evolution of *Daphnia* immunity genes: polymorphism in a gram-negative binding protein gene and an alpha-2-macroglobulin gene. *J. Mol. Evol.* **59**:498–506.

- Loker, E. S., Adema , C. M., Zhang , S. M., & Kepler , T. B., 2004. Invertebrate immune systems—not homogeneous, not simple, not well understood. *Immunol. Rev.* **198**:10– 24.
- McDonald, J. H. & Kreitman , M., 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* **351**: 652-654.
- Michel, K., Kafatos, F.C., 2005. Mosquito immunity against Plasmodium. *Insect Biochem Mol Biol*, **35**: 677 - 689.
- Moreira, Luciano A. e Jacobs-Lorena, Marcelo, 2002. Mosquitos transgênicos para o controle da malária: progressos e desafios. *Bases da parasitologia medica.* **32**(4): pp. 531-536.
- Morlais, I., Poncon, N., Simard, F., Cohuet, A., Fontenille, D., 2004. Intraspecific nucleotide variation in *Anopheles gambiae*: new insights into the biology of malaria vectors. *Am. J. Trop. Med. Hyg.*, **71**(6):795-802.
- Obbard, D.J., Linton, Y.M., Jiggins, F.M., Yan, G., Little, T.J., 2007: Population genetics of Plasmodium resistance genes in *Anopheles gambiae*: no evidence for strong selection. *Mol. Ecol.*, **16**(16):3497-3510.
- Olliaro, P., 2001. Mode of action and mechanisms of resistance for antimalarial drugs. *Pharmacol. Ther.* **89**, 207-219.
- Osta, M. A., Christophides, G. K. & Kafatos, F. C., 2004. Effects of mosquito genes on *Plasmodium* development. *Science* **303**: 2030-2032.
- Parmakelis, A, M Slotman, JC Marshall, PH Awono-Ambene, C Antonion-Nkondjio, F Simard, A Caccone, JR Powell, 2008. The molecular evolution of four anti-malarial immune genes in the *Anopholes gambiae* species complex. *BMC Evolutionary Biology* **8**:79.
- Pili-Floury, S., Leulier, F., Takahashi, K., Saigo, K., Samain, E., Ueda, R., Lemaitre, B., 2004: In vivo RNA interference analysis reveals an unexpected role for GGBP1 in the

defense against Gram-positive bacterial infection in *Drosophila* adults. J. Biol. Chem., **279**:12848-12853.

- Reimer, L. J., Tripet F., Slotman, M. Spielman, A. Fondjo E. & Lanzaro, G. C., 2005. An unusual distribution of the Kdr gene among populations of *Anopheles gambiae* on the island of Bioko, Equatorial Guinea. Insect Molecular Biology **14**(6): 683–688.
- Reuter, G. & Gabius, H.J., 1999. Eukaryotic glycosylation: whim of nature or multipurpose tool? Cell. Mol. Life Sci., **55**: 368– 422.
- Rey, L., 1992. Bases da parasitologia medica. Rio de Janeiro: Guanabara Koogan, pp. 349.
- Richie, T. L. & Saul, A., 2002. Progress and challenges for malaria vaccines. Nature **415**: 694-701.
- Röhl A., 2003. Network: a program package for calculating phylogenetic networks. Fluxus Technology Ltd. (www.fluxus-engineering.com), Hamburg.
- Rozas, J., J. C. Sánchez-DelBairro, X. Messeguer & R. Rozas, 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19**: 2496-2497.
- Sackton, T.B., Haney, R.A., Rand, D.M., 2003. Cytonuclear coadaptation in *Drosophila*: disruption of cytochrome c oxidase activity in backcross genotypes. Evolution **57**: 2315–2325.
- Sackton, Timothy B., Haney, Robert A., and, David M., 2003. Cytonuclear coadaptation in *Drosophila*: disruption of cytochrome c oxidase activity in backcross genotypes. Evolution; international journal of organic evolution, **57**(10): 2315-25.
- Schluter, D. 1996. Adaptive radiation along genetic lines of least resistance. Evolution **50**: 1766–1774.
- Schneider, S., Roessli, D. & Excoffier, L., 2000. Arlequin: a software package for population genetics. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva, Geneva.

- Shahabuddin, M., Toyoshima, T., Aikawa, M. and Kaslow, D. C., 1993. Transmission-blocking activity of a chitinase inhibitor and activation of malarial parasite chitinase by mosquito protease. *Proc. Natl. Acad. Sci.*, **90**: 4266-4270.

- Simard, F., Licht, M., Besansky, N.J., Lehmann, T., 2007: Polymorphism at the defensin gene in the *Anopheles gambiae* complex: testing different selection hypotheses. *Infect. Genet. Evol.*, **7**:285-292.

- Tahar, R., Boudin, C., Thiery, I. & Bourgouin, C., 2002. Immune response of *Anopheles gambiae* to the early sporogonic stages of the human malaria parasite *Plasmodium falciparum*. *EMBO J.*, **21**: 6673-6680.

- Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**: 585-595.

- Wang, P. & Granados, R. R., 2001. Molecular structure of the peritrophic membrane (PM), identification of potential PM target sites for insect control. *Arch. Insect Biochem. Physiol.* **47**: 110-118.

- Warburg, A. & Miller, L.H., 1991. Critical stages in the development of *Plasmodium* in mosquitoes. *Parasitol. Today*, **7**:179–181.

- Warr, C.G., Vosshall, L.B., Amrein, H.O., Carlson, J.R., Gao, Q., Smith, D.P., 2000. A unified nomenclature system for the *Drosophila* odorant receptors. *Cell* **102**(2):145-146.

- Warr, E., S., Das, Dong, Y. & Dimopoulos, G., 2008. The Gram-Negative Bacteria-Binding Protein gene family: Its role in the innate immune system of *Anopheles gambiae* and in anti-*Plasmodium* defence. *Insect Molecular Biology* **17** (1): 39 – 51.

- Waterhouse, R.M., Kriventseva, E.V., Meister, S., Xi, Z., Alvarez, K.S., Bartholomay, L.C., Barillas-Mury, C., Bian, G., Blandin, S., Christensen, B.M., Dong, Y., Jiang, H., Kanost, M.R., Koutsos, A.C., Levashina, E.A., Li, J., Ligoxygakis, P., Maccallum, R.M., Mayhew, G.F., Mendes, A., Michel, K., Osta, M.A., Paskewitz, S., Shin, S.W., Vlachou, D., Wang, L., Wei, W., Zheng, L., Zou, Z., Severson, D.W., Raikhel, A.S., Kafatos, F.C., Dimopoulos,

- G., Zdobnov, E.M., Christophides, G.K., 2007. Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science* **316**: 1738–1743.
- Wright S., 1978: Evolution and the Genetics of Populations: Variability Within and Among natural Populations. *Heredity*, **86**, 641–647.
 - Zheng, L., Cornel, A.J., Wang, R., Erfle, H., Voss, H., Ansorge, W., Kafatos, F.C., Collins, F.H., 1997. Quantitative trait loci for refractoriness of *Anopheles gambiae* to *Plasmodium cynomolgi* B. *Science* **276**: 425-428.
 - <http://www.cazy.org/>, consultado em 10/11/2008
 - <http://www.cdc.gov/>, consultado em 05/09/2008
 - <http://www.ensemble.org>, consultado em 20/09/2008
 - http://www.ensembl.org/Anopheles_gambiae/, consultado em 25/09/2008
 - http://www.ensembl.org/Anopheles_gambiae/Search/Summary?species=Anopheles_gambiae;idx=;q=gambiae, consultado em 25/09/2008
 - <http://entomology.ucdavis.edu/faculty/lanzaro/publications.htm>, consultado em 25/10/2008
 - <http://www.expasy.org>, consultado em 17/10/2008
 - <http://www.dpd.cdc.gov/dpdx/HTML/Malaria.htm>, consultado em 3/09/2008
 - <http://paginas.terra.com.br/arte/mundoantigo/africa/>, consultado em 01/10/2008
 - <http://smart.embl-heidelberg.de/>, consultado em 07/10/2008

10 Anexos

Anexo 1

Tabela 15-Resultados do teste de Tajima D

	Zona codificante		Intrão		3'	
	Tajima's D	valor p	Tajima's D	valor p	Tajima's D	valor p
Angola	1,551	p> 0.10	2,006	p < 0.05	1,44	p > 0.10
Gabão	-0,146	p> 0.10	0,986	p > 0.10	-0,936	p > 0.10
Ghana-Accra	-0,201	p> 0.10	0,755	p > 0.10	0,452	p > 0.10
Ghana-Okyereko	-0,381	p > 0.10	-0,869	p > 0.10	-0,850	p > 0.10
Guiné-Bissau	-1,245	p > 0.10	-1,120	p > 0.10	n/a	n/a
Moçambique	1,036	p > 0.10	0,626	p > 0.10	1,464	p > 0.10
Tanzânia	0,124	p > 0.10	-0,319	p > 0.10	-0,793	p > 0.10
Formas M	-0,774	p > 0.10	0,568	p > 0.10	-1,041	p > 0.10
Formas S	-0,807	p > 0.10	-1,012	p > 0.10	-1,6552	0.10 > p > 0.05

Anexo 2

Tabela 16- Resultados do teste de Fu e Li

	Zona codificante				Intrão				3'			
	D	F	p	D	F	p	D	F	p	D	F	p
Angola	1,212	1,212	p> 0.10	1,467	1,675	0.10> p> 0.05	1,052	1,158	p > 0.10	1,052	1,158	p > 0.10
Gabão-Dienga	-0,338	-0,327	p > 0.10	0,840	0,962	p > 0.10	-0,725	-0,860	p > 0.10	-0,725	-0,860	p > 0.10
Ghana-Accra	-0,243	-0,258	p > 0.10	0,389	0,507	p > 0.10	0,951	0,914	p > 0.10	0,951	0,914	p > 0.10
Ghana-Okyereko	-0,253	-0,327	p > 0.10	-1,320	-1,366	p > 0.10	-0,374	-0,555	p > 0.10	-0,374	-0,555	p > 0.10
Guine-Bissau	-1,541	-1,692	p > 0.10	-0,57	-0,825	p > 0.10	n/a	n/a	n/a	n/a	n/a	n/a
Moçambique	1,033	1,167	p > 0.10	1,239	1,219	p > 0.10	0,804	1,069	p > 0.10	0,804	1,069	p > 0.10
Tanzânia	-0,041	-0,004	p > 0.10	-0,069	-0,135	p > 0.10	-0,969	-1,010	p > 0.10	-0,969	-1,010	p > 0.10
Formas M	-1,400	-1,452	p > 0.10	1,166	1,152	p > 0.10	-0,189	-0,493	p > 0.10	-0,189	-0,493	p > 0.10
Formas S	-1,507	-1,501	p> 0.10	-1,958	-1,952	p > 0.10	0,069	-0,527	p > 0.10	0,069	-0,527	p > 0.10

Anexo 3

Tabela 17- Resultados do teste de McDonald & Kreitman

	Codificante				McDonald & Kreitman				Intrão			
	Polimorfismos		Fixas		Teste de fisher		Polimorfismos		Fixas		Teste de fisher	
	ss	ns	ss	ns	ss	ns	ss	ns	ss	ns	ss	ns
ANGOLA -Gabao	46	5	2	0	1	2	0	0	0	0	0	n/a
ANGOLA- Ghana Accra	48	4	0	0	n/a	4	0	0	0	0	0	n/a
ANGOLA- Ghana Okyereko	60	6	0	0	n/a	4	0	0	0	0	0	n/a
ANGOLA- Guine Bissau	75	10	0	0	n/a	3	0	0	0	0	0	n/a
ANGOLA- Moçambique	33	6	1	0	1	4	0	0	0	0	0	n/a
ANGOLA- Tanzânia	48	4	1	0	1	4	0	0	0	0	0	n/a
GABAO- Ghana Accra	53	7	0	0	n/a	0	0	0	0	0	0	n/a
GABAO- Ghana Okyereko	68	9	0	0	n/a	2	0	0	0	0	0	n/a
GABAO- Guine Bissau	87	13	0	0	n/a	0	0	0	0	0	0	n/a
GABAO- Moçambique	43	7	0	0	n/a	0	0	0	0	0	0	n/a
GABAO- Tanzânia	53	7	0	0	n/a	0	0	0	0	0	0	n/a
GHANA ACCRA - Ghana Okyereko	66	7	0	0	n/a	4	0	0	0	0	0	n/a
GHANA ACCRA - Guine Bissau	86	11	0	0	n/a	1	0	0	0	0	0	n/a
GHANA ACCRA - Moçambique	42	8	0	0	n/a	2	0	0	0	0	0	n/a
GHANA ACCRA - Tanzânia	56	6	0	0	n/a	3	0	0	0	0	0	n/a
GHANA OKYEREKO- Guine Bissau	83	12	0	0	n/a	3	0	0	0	0	0	n/a
GHANA OKYEREKO- Moçambique	59	10	0	0	n/a	4	0	0	0	0	0	n/a
GHANA OKYEREKO- Tanzânia	68	8	0	0	n/a	4	0	0	0	0	0	n/a
GUINE BISSAU- Moçambique	77	14	0	0	n/a	1	0	0	0	0	0	n/a
GUINE BISSAU- Tanzânia	84	12	0	0	n/a	3	0	0	0	0	0	n/a
MOÇAMBIQUE - Tanzânia	38	8	0	0	n/a	2	0	0	0	0	0	n/a
Formas M-S	105	20	0	0	n/a	2	0	0	0	0	0	n/a

ss- substituições sinónimas ns- substituições não-sinónimas n/a- sem resultados

Anexo 4

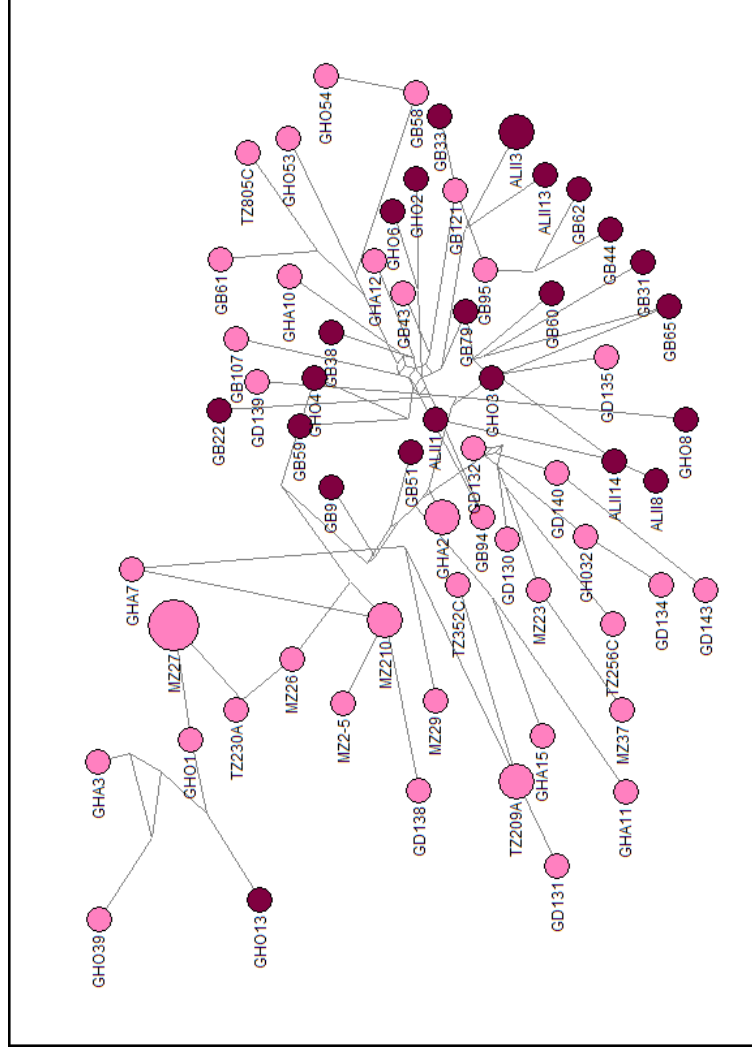


Figura 15- Rede de haplotipos referente às formas moleculares

O rosa claro corresponde à forma molecular S e o rosa escuro à forma molecular M.

Anexo 5

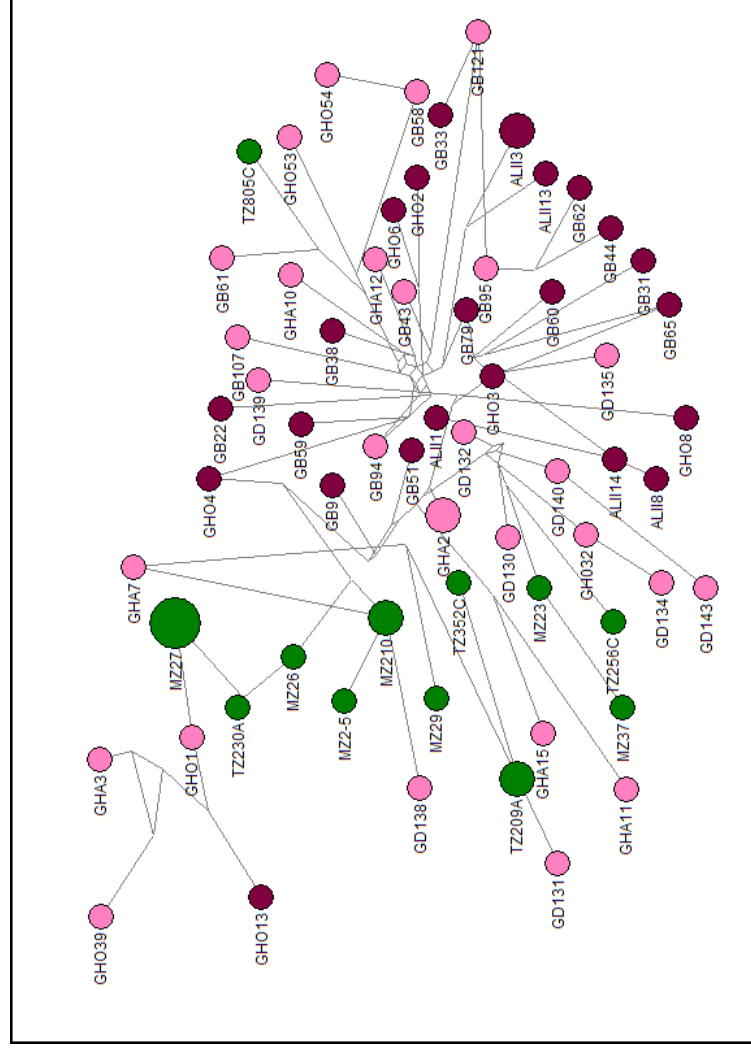


Figura 16- Rede de haplotipos referente às formas moleculares e Ocidente versus Oriente

O rosa claro corresponde à forma molecular S da zona Ocidental; a verde a forma molecular S da Zona Oriental e o rosa escuro à forma molecular M

